

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 06:15:54 ; Search time 3276.58 Seconds
(without alignments)
2615.705 Million cell updates/sec

Title: US-09-319-156a-6
Perfect score: 635
Sequence: 1 ccctgtacttcaactcctc.....tgaaaaaaaaaaaaaaa 635

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	485.6	76.5	771	10	BI087886
2	450.8	71.0	689	12	AG121669
3	437	68.8	494	9	AA781423
4	432.4	68.1	522	9	AW971553
5	423	66.6	653	12	AG033781
6	415.8	65.5	490	9	AI598135
7	397.4	62.6	440	10	BE732673
8	396	62.4	543	10	BI963185
9	394.8	62.2	651	12	AG058970
10	393.2	61.9	701	12	AG126669
11	393	61.9	609	12	AG066901
12	392.6	61.8	443	9	AA837267
13	387.2	60.9	712	12	AO892947
14	386.8	60.9	436	9	AI128526
15	385.2	60.7	446	9	AI393478
16	384.6	60.6	641	12	AG036829
17	384.2	60.5	485	9	AW511366

C 18	375.2	59.1	556	9	AU158595	AU158595
C 19	365.8	57.6	470	9	AI074704	AI074704 ox83d05.s
C 20	364.8	57.4	679	12	AG076758	AG076758 pan trogl
C 21	364.2	57.4	415	9	AI128496	AI128496 qe1e08.x
C 22	360.6	56.8	777	12	AG030228	AG030228 pan trogl
C 23	359.8	56.7	425	9	AI570707	AI570707 tm79g09.x
C 24	359.2	56.6	431	9	AA552941	AA552941 nk61a10.s
C 25	359.2	56.6	490	9	AA426511	AA426511 zw02e05.r
C 26	359.2	56.6	619	12	AG133542	AG133542 pan trogl
C 27	359.2	56.6	674	12	AG091649	AG091649 pan trogl
C 28	355.4	56.0	701	12	AG096321	AG096321 pan trogl
C 29	355	55.9	458	10	R76086	R76086 y171b03.s1
C 30	354.8	55.9	422	10	N53177	N53177 yv56h11.s1
C 31	354.6	55.8	654	12	AG063042	AG063042 pan trogl
C 32	353.8	55.7	665	12	AG088104	AG088104 pan trogl
C 33	351.2	55.3	621	12	AG098991	AG098991 pan trogl
C 34	351	55.3	433	9	AI379210	AI379210 t401q11.x
C 35	350.8	55.2	410	9	AA250958	AA250958 z507d10.s
C 36	350.6	55.2	1169	9	BM472049	BM472049 AGENCOURT
C 37	347.6	54.7	424	10	R27412	R27412 yh46d11.s1
C 38	341.2	53.7	446	9	AI288235	AI288235 q180a08.x
C 39	332.8	52.4	687	12	AG065253	AG065253 pan trogl
C 40	331.8	52.3	692	12	AG040467	AG040467 pan trogl
C 41	331.6	52.2	388	10	H01325	H01325 y199e01.s1
C 42	331.2	52.2	471	9	AA709471	AA709471 z491h06.s
C 43	330	52.0	569	12	AO580978	AO580978 RPCI-11-4
C 44	328.8	51.8	438	10	R77278	R77278 y175d06.s1
C 45	318.8	50.2	840	12	AO893377	AO893377 HS-3072.A

ALIGNMENTS

RESULT 1
BI087886 771 bp mRNA linear EST 20-JUN-2001
LOCUS 602852690F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:493894 5',
DEFINITION mRNA sequence.
ACCESSION BI087886
VERSION BI087886.1 GI:14506216
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC http://mgc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strauberg, Ph.D.
COMMENT Email: c9aps-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/MLN at:
http://image.llnl.gov
Plate: LHAM11015 row: d column: 23
High quality sequence stop: 762.
location/Qualifiers
1. 771
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:493894"
/clone_lib="NIH_MGC_10"
/cell_line="MGC36"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site:1; NotI:
Site:2; SalI: Cloned unidirectionally. Primer: oligo dT.
Average insert size 1.5 kb. Library prepared by Life
Technologies."

FEATURES

source
BASE COUNT 225 a 206 c 166 g 174 t

ORIGIN

Query Match 76.5%; Score 485.6; DB 10; Length 771;
 Best Local Similarity 89.9%; Pred. No. 7.7e-93;
 Matches 570; Conservative 0; Mismatches 49; Indels 15; Gaps 4;

OY 1 cccgtatccttaaccctccttgtaagttgtctctccagaatacaaaactgtaacta 60
 |||||
 DB 151 cccgtatccttaaccctccttgtaagttgtctctccagaatacaaaactgtaacta 210
 |||||
 OY 61 caaatgtcttcaaatgagacacagatgagtcacatgacataagatccacgtgacc 120
 |||||
 DB 211 -----CAATGAGCCCAAGATGACGCTCAAGACATCAAGATCTACCGGACGCC 258
 |||||
 OY 121 ctgagacgagcctgctagaccatgctcgcagttaatgacatgagacacccctccgag 180
 |||||
 DB 259 CTGAGACGGCCCTGCTAGCCACGATCTGATGTATATGACATCAAGGCCACCCCTCTGAG 318
 |||||
 DB 181 gaatactcaactgcaaacacccctactatgcccacatcagcgggaagacagttagagcgt 240
 |||||
 DB 319 GAAATCTAGCTGACCAACCTCTACTAGCCCAATTCAGCAGGAAGCAGTTAGAGCGGT 378
 |||||
 OY 241 catcagccaacctcccccacacgcaacttggtttctcgt-tgagagggggagacataaga 299
 |||||
 DB 379 CGTC-GGCAACCTCCCAAAAGCACTTAGGTTTCTGTGTAGATGGGGAGCTAGAGA 437
 |||||
 OY 300 cagagactagctggaattcctagagccaagaaatccctaaagccagctgaggaagtgac 359
 |||||
 DB 438 CAGGACTAGCTGAGATTCTCTAGCGTGAAGTAATCCCTAAGCCCTAGCTGGGAAGCTGAC 497
 |||||
 OY 360 tgcatacacccttaaacatgagggcttgcaacttagctcacaccgacacatcagaagct 419
 |||||
 DB 498 CACATCCACCTTTAAACAGGGGCTTGGACACTTACCTCACACCTGACCAATCAGAGACT 557
 |||||
 OY 420 cactaataatgcctaataatagcaaaatagaggta-aagaataagcaactcactctatgcc 478
 |||||
 DB 558 CACTTAATATGCTAATTAAGCAAGACAGAGAGTACAAAGAAATAGCAATCATATTTGCC 617
 |||||
 OY 479 tgaagacacagcggagagacaagatcgagatatlaaacacccagcatcagacggacgac 538
 |||||
 DB 618 TGAGAGCAGCAGAGAGAGGAGACATATGCGGATATTAAACCAAGCTTCGAGCGCGAAC 677
 |||||
 OY 539 ggcacaccccttggtgctccctcctctgtatggagcgtcgtttcactatctaacct 598
 |||||
 DB 678 GGCACACCCCTTTGGGTGCTCCCTCTGTGTATGGAGCTGTATTCATTTTCACT 737
 |||||
 DB 599 ctatcaaatctgcaactgtaaaaaaataaaaaa 632
 |||||
 DB 738 CTATTAAATCTTGCACACTGCAAAAACCAAAACA 771
 |||||

RESULT 2
 AGI21669/c 689 bp DNA linear GSS 04-NOV-2001
 LOCUS Pan troglodytes DNA, clone: PTB-130M15.F, genomic survey sequence.
 DEFINITION AGI21669
 ACCESSION AGI21669.1 GI:16650834
 VERSION GSS: GSS (genome survey sequence).
 KEYWORDS Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male
 SOURCE BAC library clone: PTB-130M15.F.
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Pan.
 1 (sites)
 Fujiiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
 Totoki, Y., Watanabe, H. and Sakaki, Y.
 TITLE BAC end sequences of library PTB
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 689)
 AUTHORS Fujiiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
 Totoki, Y., Watanabe, H. and Sakaki, Y.
 TITLE Direct Submission

JOURNAL

Submitted (02-AUG-2001) Asao Fujiiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail:chimbeseqsc.riken.go.jp, URL:http://hqp-gsc.riken.go.jp/,
 Tel:81-45-503-9111, Fax:81-45-503-9170)
 Clones are derived from the chimpanzee BAC library PTB This BAC end
 was generated during the Rad process and may have higher chance of
 clone tracking errors.
 PRIMERS
 Sequencing: -21M13

FEATURES

source
 1. 689
 Location/Qualifiers
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="PTB-130M15.F"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_11b="PTB Chimpanzee Male BAC library"

BASE COUNT 164 a 159 c 176 g 189 t 1 others
 ORIGIN

Query Match 71.0%; Score 450.8; DB 12; Length 689;
 Best Local Similarity 88.3%; Pred. No. 1.8e-85;
 Matches 515; Conservative 0; Mismatches 58; Indels 10; Gaps 2;

OY 1 cccgtatccttaaccctccttgtaagttgtctctccagaatacaaaactgtaacta 60
 |||||
 DB 666 CCNTGATCTTTAAAGCTCTGTGTAAGCTGTCTCTCAGAAATGAAAGCTTAATAACCA 607
 |||||
 OY 61 caaatgtcttcaaatgagacacagatgagtcacatgacataagatccacgtgacc 120
 |||||
 DB 606 TTAATGTTTCTCAATGAGAGCACTTATGACGTCACTAATGATCTACCATGAGGCC 547
 |||||
 OY 121 ctgagacgagcctgctagaccatgctcgcagttaatgacatgagacacccctccgag 180
 |||||
 DB 546 CTGAGACGCGCTGCTAGCCACGATGCTGTATGATGACATGAGAGGCCACCCCTCTGAG 487
 |||||
 OY 181 gaatactcaactgcaac-aacccctactatgcccacatcagcgggaagacagttagagcgt 239
 |||||
 DB 486 GAAATCTCACTGCAACAAACCTTACTACACTCCAGTTACAGCAGAAACAGATTAGAGCGG 427
 |||||
 OY 240 tcatcagccaacctcccccacacgcaacttggtttctcgttgagagggggagacataaga 299
 |||||
 DB 426 TCGTGCGCCAACTCCCAATGCGCACTTGCGTTTCTGTGTAGAGGGGAGCTAGAGA 367
 |||||
 OY 300 cagagactagctggaattcctagagccaagaaatccctaaagccagctgaggaagtgac 359
 |||||
 DB 366 CAGGACTATGCTGAGATTCTCTAGCGGCACTAAGAAATCCCTAAGCCCTAGCTGGGAAGCTGAC 307
 |||||
 OY 360 tgcatacacccttaaacatgagggcttgcaacttagctcacaccgacacatc----- 411
 |||||
 DB 306 CACATTCACCTTTAAACAGAGGCTTGGCACTTACTACACCCCAACCAATCAGGTAGTA 247
 |||||
 OY 412 -agaagctcactaataatgactaataatgaggaataatgaggaataatgacatc 470
 |||||
 DB 246 AAGAGGCTCTCAATTAATGCTTAATTAAGCAAAAGAGAGGTAAGAATAGCCATTAAT 187
 |||||
 OY 471 ctatgcttgagagacagcggagagacaagatcgagatatlaaacacccagcatcagag 530
 |||||
 DB 186 TTATTGCTTGAGAGTACAGCGGAGAGCAATGATGCGGATATAACATCAGTCATGAG 127
 |||||
 OY 531 ccggaacgcaaacaccccttggtgctccctcctctgtatggg 573
 |||||
 DB 126 CCACCAATGCTACCCCTTTGGGTGCTCCCTCTCTTTGTATGG 84
 |||||

RESULT 3
 AA781423/c

Oy	335	ccctaagccctagcttgggaaggtgactgtgacatccactctcaacaatggggcttccaacttag	394
Db	282	cccttaaaccttagcttgggaaggtgaccgcattccaccttttaacacggggcttgcacatttag	223
Oy	395	ctcacaccgcagcaaatcagagagctcactctaaatgctctaatctaggcaaaaaatgagagttaa	454
Db	222	ctcacaccgcagcaaatcagagagctcactctaaatgctctaatgctctaatgagcaaaaaatgagaggtta	163
Oy	455	agaaatagccaatctcatctatctgctcgagagacacagcgagagagaaagaatcgagatatata	514
Db	162	acaaatagccaatctcttcttctgctcgagagacacagtgagagggacaaagattgcacattata	103
Oy	515	aaaccagagcatctgagccgagcaagcgcaaacccctctgggtccccctccctctgtatlgagc	574
Db	102	aaaccagagcatctgagccgagcaaacggcaaacccctgggtgggtcccttcccttggatgagga	43
Oy	575	gctctgtttctacatctatattctacatattaaatcttgcaact	616
	42	gctctgtttttacgtctattttacgtctattttaaattgtgcaact	1

RESULT	5				
AG033781/c					
LOCUS	AC033781	653 bp	DNA	linear	GSS 01-NOV-2001
DEFINITION	Pan troglodytes DNA, clone: PTB-008021.f,				genomic survey sequence.
ACCESSION	AC033781				
VERSION	AC033781.1	GI:16560654			
KEYWORDS	GSS; GSS (genome survey sequence);				
SOURCE	Pan troglodytes male lymphoblast DNA, clone_1lb:PTB Chimpanzee Male				

ORGANISM	REFERENCE	TITLE	JOURNAL	REFERENCE	AUTHORS
Pan troglodytes					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.					
1 (sites)					
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,					
Totoki, Y., Watanabe, H. and Sakaki, Y.					
BAC end sequences of library PTB					
Unpublished					
2 (bases 1 to 653)					
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,					

JOURNAL
Submitted (02-AUG-2001) Aseo Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suchojo-cho, Tsukuba, Ibaraki, 305-0856, Japan
(E-mail: Chihimides@sc.riken.go.jp, Yokohama.kanagawa.riken.go.jp,
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the chimpanzee BAC library PB7. This BAC
was generated during the R&D process and may have higher chance of
clone tracking errors.

```

Sequencing: -21M13
LIBRARY
  Vector      : pKSI45
  R.Site 1    : SacI
  R.Site 2    : SacI
  Location/Qualifiers
    1..653
FEATURES
  source

```

```

/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-008021.f"
/sex="male"
/cell-type="lymphoblast"
/clone_lib="PTB chimpanzee Male BAC library"
BASE COUNT      145 a      155 c      178 g      173 t
ORIGIN

```

Query Match	66.6%	Score 423, DB 12	Length 653,
Best Local Similarity	88.0%	Pred. No. 1.3e-79,	
Matches 511, Conservative	0;	Mismatches 52;	Indels 18; Gaps 4;

Oy 2 cctgattctttaaaccctc-ttgttaagttttctctcccaagaaaccaaacctglaaacta 60
 Db 639 CCTGTTTCTTTTAACCCCGCTGTTAAGTGTGTCTCTCNCCAGATCATNAGCTGTAAAGCTA 580
 Oy 61 caaatgtctctcaaatgtagacaccaga tggatgcatagactaaagatccaccgttgacc 120
 Db 579 CAAATGGTCTTTCAAAATGGAGGCCCAAAATVG-----CAGTCCATGAGACC 535
 Oy 121 ctggaacggcctgtctagcccatgtctcgatgtttaaagcaatgaaagcaccctccgag 180
 Db 534 CTGGACCGCGCTCGTAGCGCCATGCTCTATGTATTAATGACATCGAAGGC -GCCCTCTTAG 476
 Oy 181 gaactctaactcgcaaacccctactatcgatcgcaacttaagagggaagagctttagaggt 240
 Db 475 AAAATCTCAACTGCACAACCCCTACTATAACCCCGATTGAGAGGAAGAGATTAAGTGT 416
 Oy 241 catcagccaacctccccaacaagcaacttggtttctcgttttagaaggggagactagagac 300
 Db 415 CGTTGGCCCAACTCCCCACAACGACTTGGGTTTCTCTTGTAGAGGGGGACTAGAGAC 356
 Oy 301 aggaactagctgatttctcttagcgcaagcaagaatccctaagccttagctggaagtgact 360
 Db 355 AGGACTTACCTGGATTTCCTAAGCCAACTAAGATCCTTAAGCCTCTGGAAGGTGACC 296
 Oy 361 gcatccaccctctaaacaatggygctctgcaacttagctacacccgaccaatca -gagagct 419
 Db 295 ACATTCACACCTTTAAACAATGGGGCTTGGCAACTTAGCTCACACCACAACGTAAGGAGACT 236
 Oy 420 caactaaatgtctaacttagcgcaaaaataagaggtlaagaaatagccaactcatctagtct 479
 Db 235 CACTTAAATGCTAATTAGGCAAAAAACAGAGAGTAAAGAAATAGCCAAATCATCTATTGCT 176
 Oy 480 gagagcacagcgaggagcaagaagatcgggatataaaccagcaactctgagccggagcg 539
 Db 175 GAGAGCACAGCAGAGAAGACAAATGATGCAGATATTAACCCCAAGCATTTGAGCAGCAATG 116
 Oy 540 gcaacccctctgggtccctctcctctttagttagggcgctcg 580
 Db 115 GCTACCTTCTTTGGGTCCCTCCCTTTTATAGGGAGCCG 75

RESULT	6
A1598135/c	
LOCUS	
DEFINITION	490 bp mRNA linear EST 12-MAY-1999
	tnt4a10.x1 NCI CGAP Brn25 Homo sapiens cDNA clone IMAGE:2167578 3'
	similar to contains PTR5.t1 PTR5 repetitive element ;, mRNA
	sequence.

ACCESSION	AI598135	GI:4607183
VERSION	AI598135.1	
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	

REFERENCE AUTHORS TITLE	JOURNAL COMMENT
<p>Eumetazoa: Metazoa: Chordata: Craniata: Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.</p> <p>NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.</p> <p>National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BIGAP), Tumor Gene Index</p>	<p>Unpublished (1998)</p> <p>Contact: Robert Strausberg, Ph.D.</p>

Email: cgabos-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-GAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrp/image/image.html
Insert length: 1407 Std Error: 0.00
Seq primer: -40UP from Gluco

DEFINITION	1657c10.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
ACCESSION	U00001.1
VERSION	1
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 543)
AUTHORS	Melton,D., Brown,J., Kenly,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Scarce,M., Bressell,J., Gradwohl,G., Clifton,S., Hillier,T., Marra,M., Page,D., Wylie,T., Martin,J., Blisrah,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagarelshvili,R., Williams,T., Jackson,Y. and Bowers,Y.
TITLE	Endocrine Pancreas Consortium
JOURNAL	Unpublished (2000)
COMMENT	Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue Endocrine Pancreas Consortium Harvard University, Howard Hughes Medical Institute Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138 Tel: 617-495-1812 Fax: 617-495-8557 Email: dmelton@bldp.harvard.edu Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Juliana Brown (brown@fas.harvard.edu) High quality sequence stop: 451. Location/Qualifiers 1..543 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="Melton Normalized Human Islet 4 N4-HIS 1" /sex="both" /tissue_type="Islets of Langerhans" /dev_stage="Adult" /lab_host="DH10B" /note="Organ: Pancreas; Vector: pSPORN1; Site.1: Not 1; Site.2: Sal 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an Ecot 05 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."
BASE COUNT	116 a 123 c 147 g 156 t 1 others
ORIGIN	
Query Match	62.4%; Score 396; DB 10; Length 543;
Best Local Similarity	88.3%; Pred. No. 7.2e-74;
Matches 467; Conservative	0; Mismatches 51; Indels 11; Gaps 3;
Db	543
Db	483
Db	194
Db	424

RESULT	9	AG058970/c
LOCUS	AG058970	
DEFINITION	Pan troglodytes DNA, clone: PTB-046A08.R, genomic survey sequence.	
ACCESSION	AG058970	
VERSION	AG058970.1	GI:16596431
KEYWORDS	GS: GS (genome survey sequence).	
SOURCE	Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male BAC library clone:PTB-046A08.R.	
ORGANISM	Pan troglodytes	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.	
AUTHORS	1 (sites)	
TITLE	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Tokoki, Y., Watanabe, H. and Sakaki, Y.	
REFERENCE	BAC end sequences of library PTB	
AUTHORS	Unpublished	
TITLE	2 (bases 1 to 651)	
JOURNAL	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Tokoki, Y., Watanabe, H. and Sakaki, Y.	
REFERENCE	Submitted (02-AUG-2001) Ageo Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Shuhiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: fujibae@gscc.riken.go.jp, URL: http://npg.gscc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)	
COMMENT	Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.	
PRIMERS		
SEQUENCING	M13Rev	
LIBRARY		
VECTOR	: pKS145	
R.Site 1	: SacI	
R.Site 2	: SacI.	
Location/Qualifiers		
1..651		
/organism="Pan troglodytes"		
/db_xref="taxon:9598"		
/clone="PTB-046A08.R"		
/sex="male"		
/cell_type="lymphoblast"		
/clone_11b="PTB Chimpanzee Male BAC Library"		
BASE COUNT	141 a 160 c 159 g 191 t	
ORIGIN		

Query Match	Similarity	62.2%	Score 394.8;	DB 12;	Length 651;
Best Local	Similarity	87.6%	Prod. No. 1.2e-73;		
Matches 503;	Conservative	0;	Mismatches 57;	Indels 14;	Gaps 6;
QY	20	ttgttaagttgtctcctcca-gaatcaaacctttaaactcaaatgtgtcttccaatg	78		
Db	613	TTGATAGGTTTGTCCTCTCCCAAGAAATCAAAAGCTGTAAACTATCAAAATGTTCTTCAATG	554		
QY	79	gagcagcagatgaggtccatgac-taagatccacgttgaacccctgagccgctgtag	137		
Db	553	GAGCACAGATACAGTCATGCTTAAGTCTCCCGTGAACCCCTGGACGTGGCAGGTAG	494		
QY	138	cccatgtcccgatgttaatgacatgtgaagcaacccctccgagaatcatctaacgtca	197		
Db	493	CCCAATGCTCTGATGTTAATGACATTTGAAGGACACCCCTCCGAGGAAATCTCACTGCACA	434		
QY	198	accctactatgcccacatcagcgggaagcagttagacggtcaacgaacactccc	257		
Db	433	ACCCTGCTACACCCCAATTCAGCAGGAACGATTAGACAGTCTCAGCCAACTCCCC	374		
QY	258	aacagcacttgggttttcctgttgaagggggagcgtgagacagagactgtgatttc	317		
Db	373	AACAGCACTTGGGTTTCTCTGTTGAGAGTGGGAGATGAGACAGAACTACTGGATTTC	314		
QY	318	ctagggcaacgaagaatcccttaagcctagct-gggaaagtactgtcatcccttaac	376		
Db	313	CTAGGCTCACTAAAGATTTCTTAAGGCTAGCTGGGGAAGTATGTCACACCTTTAAMC	254		
QY	377	atggagcttgacacttagctcaaccgcgaacatag-----agaagtaactaaa	427		
Db	253	ATGGGGCTTGTACCTCACCTCACCCCAAAATAGTACTAAAAAGGCTCACTTTAAA	194		
QY	428	tgctaatatagc-aaaaatagagggtlaaagaatatg-ccaatcatctatctgcctaga	485		
Db	193	TACAAATTAGCTTAACACGAGAGATTAAGAATGTCAAAATCATATATCACTGAGACC	134		
QY	486	acagcggagaggaacgaagatcgagatataaacccaagcattcgagccggcaacgcaac	545		
Db	133	ACAGTGGGAGGACATATGTCGGGTATTAAMCCAGCATTCGAGCAGAGTGCACACC	74		
QY	546	cccttgggtccccccttggatgggcgtct	579		
Db	73	CGCTTGGGTCCCTCCCATTTGTATGGAGGCTCT	40		
RESULT 10					
AG126669/c					
LOCUS	AG126669	701 bp	DNA	linear	GSS 04-NOV-2001
DEFINITION	Pan troglodytes DNA, clone: PTB-137E19.F, genomic survey sequence.				
ACCESSION	AG126669				
VERSION	AG126669.1	GI:16655834			
KEYWORDS	GSS; GSS (genome survey sequence).				
SOURCE	Pan troglodytes male lymphoblast DNA, clone_1lb:PRB Chimpanzee Male BAC library clone:PTB-137E19.F.				
ORGANISM	Pan troglodytes				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	1 (sites)				
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Pan.				
TITLE	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,				
JOURNAL	Totoki, Y., Matanabe, H., and Sakaki, Y.				
REFERENCE	BAC end sequences of library PTB				
AUTHORS	2 (bases 1 to 701)				
	unpublished				
TITLE	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,				
JOURNAL	Totoki, Y., Matanabe, H., and Sakaki, Y.				
REFERENCE	BAC end sequences of library PTB				
AUTHORS	2 (bases 1 to 701)				
	unpublished				
TITLE	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,				
JOURNAL	Totoki, Y., Matanabe, H., and Sakaki, Y.				
REFERENCE	BAC end sequences of library PTB				
AUTHORS	2 (bases 1 to 701)				
	unpublished				
TITLE	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,				
JOURNAL	Totoki, Y., Matanabe, H., and Sakaki, Y.				
REFERENCE	BAC end sequences of library PTB				
AUTHORS	2 (bases 1 to 701)				
	unpublished				
TITLE	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,				
JOURNAL	Totoki, Y., Matanabe, H., and Sakaki, Y.				
REFERENCE	BAC end sequences of library PTB				
AUTHORS	2 (bases 1 to 701)				
	unpublished				
TITLE	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,				
JOURNAL	Totoki, Y., Matanabe, H., and Sakaki, Y.				
REFERENCE	BAC end sequences of library PTB				
AUTHORS	2 (bases 1 to 701)				
	unpublished				
TITLE	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,				
JOURNAL	Totoki, Y., Matanabe, H., and Sakaki, Y.				

FEATURES		source	
Clone tracking errors. PRIMERS Sequencing: -21M13 LIBRARY Vector : pRS145 R.Site 1 : SacI R.Site 2 : SacI. Location/Qualifiers 1. .701 /organism="Pan troglodytes" /db_xref="taxon:9598" /clone="PTB-137E19.F" /sex="male" /cell_type="Lymphoblast" /clone_11b="PTB Chimpanzee Male BAC library" BASE COUNT 162 a 162 c 182 g 195 t ORIGIN			
Query Match	61.9%	Score 393.2;	DB 12; Length 701;
Best Local Similarity	88.1%;	Pred. No. 2.6e-73;	
Matches 465;	Conservative 0;	Mismatches 53;	Indels 10; Gaps 3;
QY 79	gagcaccagatgagatgccataagatcacccgttgagacccttgacgcgcctctagc	138	
DB 651	GACCCCGAGATCCATCCATGACTGAAGACTGCTGGAACCCCTGACAGGCTCTCTAC	592	
QY 139	catctccgatgttaatgatcatgaagagaccctc-ccgaggaatctcaatgcaca	197	
DB 591	CCATCTGCTCATGTTAATGACATTGAAGACCCTCGCCGAGAAATTCACATGCACA	532	
QY 198	accctactatgcccaattcagcgggaagcagtagagcgtlcalcaagcaacctcccc	257	
DB 531	ACCCCTACTATGCCCCGATTCAGCAGAGGACAGTAAAGTGTCGTGGCCAAATCCCC	472	
QY 258	aacagcacttggtttctcgttgagagggggagctgagagacagagactctgatttc	317	
DB 471	AACACAGTGTGTTTCTGTTGAGAGGGGACTGAGAGACAGACACTGATGCTTC	412	
QY 318	ctagagcaacgaagaatcccttaagcttagcttggaaggtgactcatcactctaaaca	377	
DB 411	CTAGGCTGACTAAGANTCCCTTAAGACTAGCTGGAGAGGTGACACACTTCACCTTGAAACA	352	
QY 378	tggggtctgcaacttagctcacaccgcgaacaatc-----agagagctcaactaaatg	429	
DB 351	CGGGGCTTGCAACTTAGCTACATCCGACATCCAGTAAATGAAGAAATTAACCTAAATG	292	
QY 430	ctaattaggcaaaataggaggtlaaagaatatgccaatcatctatctgctgagagacag	489	
DB 291	CTAATTAGCAAAACAGGAGGTAAAGAAATAGCCATATCATCACTTGAGATCAG	232	
QY 490	cgaggaggagcaagatccgggataataaacccaaagcatctagagccgggaaggaaccccc	549	
DB 231	CAGGAGGGACATGATCGGGATATTAACCCAGGACATTCAGGACGACATACAGCTACCTCT	172	
QY 550	ttaggttccctccctctgtatgagcgctctgtttcactctatttcac 597		
DB 171	TTTGGTCCCTCCCTTTGTATGGAGGCTGTGTTTCACTATTAAC 125		
RESULT 11	AG066901	609 bp	DNA linear GSS 03-NOV-2000
LOCUS	AG066901/C		
DEFINITION	Pan troglodytes DNA, clone: PTB-056L15.F, genomic survey sequence.		
ACCESSION	AG066901		
VERSION	AG066901.1	GI:16618703	
KEYWORDS	GSS; GSS (genome survey sequence).		
SOURCE	Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male BAC library clone: PTB-056L15.F.		
ORGANISM	Pan troglodytes		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.		
REFERENCE	1 (sites)		

AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
 Tozaki, Y., Watanabe, H. and Sakaki, Y.
 TITLE BAC end sequences of library PTB
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 609)
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
 Tozaki, Y., Watanabe, H. and Sakaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suenho-chou, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail: schimpesgsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 Clones are derived from the chimpanzee BAC library PTB this BAC end
 was generated during the Rad process and may have higher chance of
 clone tracking errors.
 COMMENT PRIMERS
 Sequencing: -21M13
 LIBRARY
 Vector : pKS145
 R.Site 1 : SacI
 R.Site 2 : SacI.
 FEATURES
 source 1. 609
 Location/Qualifiers
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="PTB-056L15.F"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_lib="PTB Chimpanzee Male BAC library"
 BASE COUNT 127 a 141 c 165 g 176 t
 ORIGIN
 Query Match 61.9%; Score 393; DB 12; Length 609;
 Best Local Similarity 87.7%; Pred. No. 3e-73;
 Matches 478; Conservative 0; Mismatches 55; Indels 12; Gaps 4;

Qy 576 cctcg 580
 |||
 Db 67 CTCGG 63
 RESULT 12
 AA837267/c 443 bp mRNA linear EST 31-MAR-1998
 LOCUS od26b10.s1 NCI CGAP GCBI Homo sapiens cDNA clone IMAGE:1369051
 DEFINITION similar to contatns PPR7.tl PPR7 repetitive element;; mRNA
 sequence.
 ACCESSION AA837267
 VERSION AA837267.1 GI:2912466
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 443)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs.rem@n1.nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
 Ph.D., Gerald Marti, M.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bhrp/image/html
 Insert Length: 1375 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham.
 FEATURES
 source 1. 443
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1369051"
 /clone_lib="NCI-CGAP_GCBI"
 /tissue_type="germinal center B cell"
 /lab_host="DH10B"
 /note="Vector: p773D-Pac (Pharmacia) with a modified
 polylinker. Site.1: Not I; Site.2: Eco RI; 1st strand cDNA
 was prepared from human tonsillar cells enriched for
 germinal center B cells by flow sorting (CD20+, IgD-),
 provided by Dr. Gerald Marti (CIBR). cDNA synthesis was
 primed with a Not I - oligo(dT) primer
 [5'-TGTTACCAATCTGAAAGTGGGCGCGCCATATTTTCTTTTCTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified p773 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 97 a 95 c 113 g 137 t 1 others
 ORIGIN
 Query Match 61.8%; Score 392.6; DB 9; Length 443;
 Best Local Similarity 94.1%; Pred. No. 4e-73;
 Matches 418; Conservative 0; Mismatches 25; Indels 1; Gaps 1;

QY 304 actagctgattctctagccaaagaaatccctaagcttagctgggaagtactgca 363
 Db 323 ACTAGCTGATTCTCTAGCCGAATTAAGAAATCCCTTAACCTAGTGGGAAGTGACCGCA 264
 QY 364 tcaactctaaacatggtgctgaacttagctcacacccagccaatcagagactact 423
 Db 263 TCCACTCTTAAACACGGGGCTTGCACTAGCTCACACCCAACCAATCAGAGACTACT 204
 QY 424 aaatcttaatttaggcaaaataggaggttaagaataagcaatcatcttctgtaga 483
 Db 203 AAATCTTAATTAGGCAAAACAGAGTAAGAAATAGCCAAATCATCTATTGCTTAGA 144
 QY 484 gcaacagggaggaacaaagatcgagatacaacccagagcatcgaacggcaagca 543
 Db 143 GCACAGTGGAGGACAGAGATTCATATTAACCCAGCATTCGACCCAGC-ANAGCAA 85
 QY 544 ccccttgggtccctcccttctgtagggcgctctgtttcactatctactact 603
 Db 84 CCGCTTGGGTCCCTTCCCTTGTATGGAGCTCTGTTCATCTATTCACTCTATT 25
 Db 604 aaattgcaactgaaaaaanaa 627
 Db 24 AAATCTTGCACTGAAAAAANA 1

RESULT 13
 AOB92947/c 712 bp DNA linear GSS 10-NOV-1999
 LOCUS HS_3131_B2_E04_MR CIT Approved Human Genomic Sperm Library D Homo
 DEFINITION sapiens genomic clone Plate=3131 Col=8 Row=J, DNA sequence.
 ACCESSION AOB92947
 VERSION AOB92947.1 GI:6349137
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 712)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 3131 row: J column: 8
 Seq primer: M13 Reverse
 Class: BAC ends
 High quality sequence stop: 712.

FEATURES
 source

1. 712
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_1lb="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelBAC11; BAC Clones in
 E-Coli DH10B"
 BASE COUNT 161 a 149 c 169 g 232 t 1 others
 ORIGIN

Query Match 61.0%; Score 387.2; DB 12; Length 712;

Best Local Similarity 82.2%; Pred. No. 4.8e-72;
 Matches 484; Conservative 0; Mismatches 94; Indels 11; Gaps 3;

QY 56 aactacaattgttctcaaatgagacacagatggtccatgactagaatccacgtg 115
 Db 699 AAGTTACAAAGGTTTCTNCAAGAGATCCCAAGATGCTAT-ACTCAAATCTCCGG 641
 QY 116 gaccctcgacccggtccgtcagccatgctccgaagttaagatgaagagccctc 175
 Db 640 ACCCTTGGGACCGGCTCGTATGTCATGCTTCCATGTGTGATGATTAAGAGCACCTTC 581
 QY 176 ccgaggaatctcaactgacaaacccctactatgcccnaatcagcgggaagcagtaga 235
 Db 580 CCGAGAAATCTCAAGTGAAGTACCTTAGT-TGCACACAGTTCAGAGAAAGCAGTTAGA 522
 QY 236 gcggtatcagccaaactccccaacagcaacttgggtttctctgttgaaggaggactga 295
 Db 521 GCGGCCCTTGGCCAACTCCCAATGCTTGGGTTTCTGTGAGAGGGGTTGCTGA 462
 QY 296 gagaacagactagctgatttccctagccaagagaatccctaagcttagctggagg 355
 Db 461 GAGACAGACTAGCTGATTTCTTAGGCCGACTAAGATCCCTAAGCTTAGCTGGAAGG 402
 QY 356 tgactgcatccactcaaatatggtggtctgcaacttagctcacaccgacaaat----- 410
 Db 401 TGACTGCATCCACTTAAACACGGGGCTTGCAACGTAGCTACACCCGACCAATGAGGT 342
 QY 411 -----cagaagctcactcaaatgctaattaggcgaataatggaggttaagaataagccaa 466
 Db 341 AGTAAAGAGAGCTCCTCAATGCTAATTTAGGCAAAACAGGAATGAAGAAATAGGCCAA 282
 QY 467 tcaactatgctctagaagcaacgaggaaggaacagatcgagatacaaacacagcatt 526
 Db 281 TCATCTATCACTGAGACACAGGGGAGGACATGATCAGATATTAACCCAGGGGTT 222
 QY 527 cgagcggcaacggaaccccttgggtccctccctctgtagggcgctctgttca 586
 Db 221 CTACCGGCAACGGCTCCTCTTGGGTACCCCTTGTATGGAGAGCTCTGTTTCA 162
 QY 587 ctctattcctctatcaatctctgcaactgcaacgaaaaaanaaaaaa 635
 Db 161 CTCTATTAACTTGCANAAGACACAAACCAACCAACCAACCA 113

RESULT 14
 A1128526/c 436 bp mRNA linear EST 27-OCT-1998
 LOCUS g651h10.x1 Soares_placenta_8c9weeks_2bHP8t09w Homo sapiens cDNA
 DEFINITION clone IMAGE:1714147 3' similar to contains PIR7.b1 PIR5 repetitive
 element ;, mRNA sequence.
 ACCESSION A1128526
 VERSION A1128526.1 GI:3597040
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 436)
 AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 This clone is available royalty-free through LNC; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert length: 720 Std Error: 0.00
 Seq primer: -40m13 fwd. RT from Amersham
 High quality sequence stop: 428.

FEATURES
 source

1. 436
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

OY	301	agggctagctgagattcttccttagggccaacgaagaaatccctaagctctagctgggaaggtgact	360
Db	2611	agggctagctgagattcttccttagggctgagtaagaaatccctaagctctagctgggaaggtgacc	2670
OY	361	gcatccacctcttaaaacatctgggctctgccaacttagctctcacaccggacaacatcagaagctc	420
Db	2671	acatcccccctcttaaaacacggggctctgccaacttagctctcacaccctgccaactcagaagctc	2730
OY	421	acttaaaatgctaaatttaagcgaaaaaatgagagtgaaagaataagcacaatcatctatgctctg	480
Db	2731	acttaaaatgctaaatttagcgaaaaaacagaggtgaagaagaatagccaatcatctatgctctg	2790
OY	481	agagacacagcggggggagacaaagatcgggagataaaacccaagcatctcgagccggcaagcg	540
Db	2791	agagacacagcggggggagacaaatgcatcgggagataaaacccaagctctcgagccggcaagcg	2850
OY	541	caaacccctctgggctcccccctctctgtactatggcgctcgtttctcaattatctaacct	600
	2851	caaacccctctgggctcccccctctctgtatgggagctcgtttctcaattatctaacct	2910
OY	601	atttaactctggcaacttataaaaaaaaaaaaaaaaaa	635
Db	2911	atttaactctggcaacttataaaaaaaaaaaaaaaaaa	2945

RESULT 2
 US-08-686-878A-50
 : Sequence 50. Application US/086686878A
 : Patent No. 5708157
 : GENERAL INFORMATION:
 : APPLICANT: Jacobs, Kenneth
 : APPLICANT: McCoy, John
 : APPLICANT: Lavallee, Edward
 : APPLICANT: Racie, Lisa
 : APPLICANT: Metberg, David
 : APPLICANT: Treacy, Maurice
 : APPLICANT: Evans, Cheryl
 : APPLICANT: Spaulding, Vilki
 : TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
 : TITLE OF INVENTION: ENCODING THEM
 : NUMBER OF SEQUENCES: 71
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Genetics Institute, Inc.
 : STREET: 87 Cambridgepark Drive
 : CITY: Cambridge
 : STATE: Massachusetts
 : COUNTRY: U.S.A.
 : ZIP: 02140
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/686, 878A
 : FILING DATE:
 : CLASSIFICATION: 536
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Brown, Scott A.
 : REGISTRATION NUMBER: 32,724
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (617) 498-8224
 : TELEFAX: (617) 876-5851
 : INFORMATION FOR SEQ. ID NO: 50:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 279 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: double
 : TOPOLOGY: linear
 : MOLECULE TYPE: cDNA
 : US-08-686-878A-50

	Query Match	Similarity	Best Local	Matches	Score	DB 1	Length	279;
					36.8%;			
					88.1%;	Pred. No. 7,6e-66;		
					Conservative	1;	Mismatches	32;
							Indels	0;
							Gaps	0;
Oy	358	actgcatccacccctctaaacatbaggctctgcaacttagctcacaccgcagcaatcagaag	1					417
Db	1	RCACATCATCCACTTTTAAACACCGGGNTTGCAAANAGATNMACACTTGACCAATCGAGAG	60					
Oy	418	ctcactaaabtgctaattaggcagaabaatagaggtcaagaagaatagcaatctcatbgc	477					
Db	61	NTCANTTAAATGATNATTNNGGCAAAAACAGAGGTAAAGAAATATACCATCATCTATTGC	120					
Oy	478	ctgaaagacacagcggaggaaggaaacaggaatccgggaatataaacccaagcaatcggacggcga	537					
Db	121	CTGAAGACACACAGACAGAGGAGACAAATGATCGGATTTAAACCCCAAGCTTTNNGAGCCGGCA	180					
Oy	538	cggcaacccccctctggctcccccctccctctgtatctggcgccctctgtcttcaactatctcag	597					
Db	181	CGGCAACCCCCCTTTGGTGCCCTCCCTTTGTATGGAGAGCTNTGTTTCATGCTATTTCAN	240					
Oy	598	tctatataactctgcaactctgaaaaaaaaaaaaaaaaaaa	635					
Db	241	TNTATTAAATNTTGCACCTGCAAAAAAATAAAAAA	278					

```

US-08-721-489-4
: Sequence 4, Application US/08721489
: Patent No. 5786465
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John
: APPLICANT: Lavallee, Edward
: APPLICANT: Racie, Lisa
: APPLICANT: Merberg, David
: APPLICANT: Treacy, Maurice
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
: TITLE OF INVENTION: ENCODING THEM
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 CambridgePark Drive
: CITY: Cambridge
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/721,489
: FILING DATE:
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Brown, Scott A.
: REGISTRATION NUMBER: 32,724
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 498-8224
: TELEFAX: (617) 876-5851
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 279 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: US-08-721-489-4

```


Query Match	36.8%;	Score 233.6;	DB 1;	Length 279;
Best Local Similarity	-88.1%;	Pred. NO. 7.6e-66;		
Matches 245; Conservative	1;	Mismatches 32;	Indels 0;	Gaps 0;

OY	358	actgcattccaccttaacatgagggtcgaacttgcataccagcaccaatcagaag	417
Db	1	RCCACATCCACGCTTTAAACAAGGGGGNTTGCAAAAAGATTNACGCTTACCATATAGAG	60
OY	418	cctcaactaaatgctaattaggcaaaaatlaguaggltaaagaatatgccaatcatltgc	477
Db	61	MTCANATAAAGATTNATTNGCCCCAAAAACAGAGCGTTAAAGAATGCCAACATCATTATTGC	127
OY	478	ctgagagcacagcggggggagcaagaagtgggagataataaccgagcaattcgaagccggcaa	537
Db	121	CTGAGAGCACAGCGAGGAGGGAACAATGTATCGGGATATTAACCAGTTTTCAGAGCGGCCAA	180
OY	538	cgagcaaccccccttggtgltccccttcctttglatgagggcgtctgttttctaactltaaac	597
Db	181	CGGCAACCCCCTTTGGGGTCCCTCCCTTGTGATGGGAGCTWTGTTTCATGCTATTCAN	240
OY	598	tctattaactcttgcactgaaaaaaaaaaaaaaaaaaaaa	635
Db	241	TNTATTAAATTTCACAATGCAGAAAAAAAAAAAAAAAAAAAAA	278

RESULT 4

US-08-691-563C-46
; Sequence 46, Application US/08691563C

```

1  GENERAL INFORMATION:
2  APPLICANT:  Herve PERRON
3  APPLICANT:  Frederic BESEME
4  APPLICANT:  Frederic BEDIN
5  APPLICANT:  Glauca PARANHOS-BACCALA
6  APPLICANT:  Florence KOMURIAN-PRADEL
7  APPLICANT:  Colette JOLIVET
8  APPLICANT:  Bernard MANDRAND
9  TITLE OF INVENTION:  VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
10 TITLE OF INVENTION:  ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
11 TITLE OF INVENTION:  THERAPEUTIC PURPOSES
12 NUMBER OF SEQUENCES:  92
13 CORRESPONDENCE ADDRESSES:

```

COMPUTER READABLE FORM:

```
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT ADDITION DATA:
```

APPLICATION NUMBER: US/08/091,363C
 FILING DATE: 03-AUG-1986

ATTORNEY/AGENT INFORMATION ;

REGISTRATION NUMBER: 30.024

REFERENCE/DOCKET NUMBER: WE

TELEPHONE: 703-836-6400

TELEFAX: 703-836-2787

SEQUENCE CHARACTERISTICS:

LENGTH: 1659 base pairs
TYPE: nucleic acid

```
STRANDEDNESS: single
```

MOLECULE TYPE: CDNA

US-08-691-563C-46

Query Match	20.2%	Score 128;	DB 3;	Length 1859;
Best Local Similarity	62.7%;	Pred. No. 2.2e-31;		
Matches 237; Conservative	0;	Mismatches 130;	Indels 11;	Gaps 2;

QY	1	cccgatctcttaaacccctctgttaagtctgtctctctccagaataaaccgtlaaaacta	60
Db	1446	CCCTGTTATTTTAACTCTCTTGCAAAATTTGTTCTTCTTAGATGAGGCCATCAAGCTA	1505
QY	61	caaatgtctctcaaatgagacacagatgagatccatbactaaagatccaccgttgacc	120
Db	1506	CAGATGGTCTTAACAAAATGGAACCCAAATGAGTCAATCACTTCTACTGAGACCC	1565
QY	121	ctggaaccgagccctgctagcccaatgctccgaatgtaatgatacatgaaagcaaccctccgag	180
Dc	1566	CTAACCCAAACCCCTGGCCCC-----TTTCCACGCGCTTAAAGAGTTCCCTCTTGAGAG	1616
QY	181	gaatctcacaactgacaaaccctactatgtccccaattgaagcgggaagacagttagagcgt	240
Db	1617	GACACTACCACTGCAAGGGCCCCATCTTTGGCCCCATTCCGAAGAGATGCTAGACAGT	1676
QY	241	catcagcacaacctccccaacagacacttgggtttctctgtctgagagggagactggaagac	300
Db	1677	CATTGCCCAA--TTCCCAAGACAGCTGGGGGTGCCCGTTTGAAGTGGGGATTGAGAGGT	1734
QY	301	agagactagctcgtgattctctcaggccaagaagaatccctaatgacctgctgggaagtgaact	360
Db	1735	GAAACCCAGCTGGACTTCTGGGTCGGGTGGGGACTTGGAAGACTTTTGTCTAGCTAAAG	1794
QY	361	gcacccaccctctaaact 378	
Db	1795	GATTGTAAATGCACAAAT 1812	

RESULT

US-09-078-294-4
; Sequence 4, Application US/09078294

GENERAL INFORMATION:

; APPLICANT: Du Sart, Desiree

TITLE OF INVENTION: A NOVEL, MICHAEL R. AFFELICANI: CANCELLID, MICHAEL R.

FILE REFERENCE: DAVIES COL

CURRENT FILING DATE: 1998-05-13

NUMBER OF SEQ ID NOS: 29

: SEQ ID NO 4

LENGTH: 80246
TYPE: DNA

ORGANISM: Nucleotide sequence of NC-cont

C
C
C
C
C
B
C
A

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Best Local Similarity	Pred. No.
67.28;	

```
Matches      213;  Conservative      0;  Mismatch
```

Qy 1 ccctgtatctttaacctccttgtaagtgtgc

Dh 58888 cccctatatttaaacctctatctcaaatatttt

Query Match	19.8%;	Score 125.8;	DB 4;	Length 80246;
Best Local Similarity	67.2%;	Pred. No. 9.3e-30;		
Matches 213; Conservative	0;	Mismatches 92;	Indels 12;	Gaps 2.

Oy	1	ccggtatcttaactacctctgtttaaagttgtctctctccagaatcaaaagttaaacta	60
Db	58888	cccgatattttaaacctctctgttccaattgtttctctcagaatcgaagccatcaagta	58947
Oy	61	caaatgtctcttcaaatgagagaccagatgagatccatgactaagaatccacgctgagacc	120
Db	58948	cagatgattcttaacaattgtaacccccaaatgagctcaactcaaacactctgcctgagagacc	59007
Oy	121	ctgagccgagcctgtctagcccatgctcgcgaatgtttaaagacattgaaagcaacctcccgag	180
Db	59008	ctgagaccgagcccgctgagcc-----ttccaatgacctaaagagctccctcctlgag	59058
Oy	181	gaaatctcaatcagacaacacccctactatgagccccaattctcgaggaagcgttagagcggt	240
Db	59059	gacactaccacatcgagagccctctctctcaacccctacacagacggaagtgtactacagcgtt	59118


```

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/011,398
; FILING DATE: 29 JAN 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/160002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2417
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-464-051-1

```

```

Query Match          5.2%; Score 33; DB 1; Length 2417;
Best Local Similarity 58.8%; Pred. No. 1;
Matches 57; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

```

```

Qy 539 ggcacaccccttgggtccccccttgatgagcgctctgttcaactcattcaact 598
    || || || || || || || || || || || || || || || || || || ||
Db 162 GGAAGGCCCTTGGAGACCTTACCCTGGCTGGTGTGACTTTGTATTAATAATTT 103
Qy 599 ctatlaactctgcacactgaataaaaaaaaaaaaaaa 635
    || || || || || || || || || || || || || || || || || || ||
Db 102 TAACCTACCTTAATTAATTAATAAAAAAAAAAAAAA 66

```

RESULT 11

US-08-462-498-1/C
Sequence 1, Application US/08462498
Patent No. 5852169

GENERAL INFORMATION:

APPLICANT: Roger Brent
APPLICANT: Antonis S. Zervos
TITLE OF INVENTION: MAX-INTERACTING PROTEINS AND RELATED
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,498
FILING DATE:

CLASSIFICATION:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/011,398
FILING DATE: 29 JAN 1993
ATTORNEY/AGENT INFORMATION:
NAME: Paul T. Clark
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/160001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

```

; LENGTH: 2417
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-462-498-1

```

```

Query Match          5.2%; Score 33; DB 2; Length 2417;
Best Local Similarity 58.8%; Pred. No. 1;
Matches 57; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

```

```

Qy 539 ggcacaccccttgggtccccccttgatgagcgctctgttcaactcattcaact 598
    || || || || || || || || || || || || || || || || || || ||
Db 162 GGAAGGCCCTTGGAGACCTTACCCTGGCTGGTGTGACTTTGTATTAATAATTT 103
Qy 599 ctatlaactctgcacactgaataaaaaaaaaaaaaaa 635
    || || || || || || || || || || || || || || || || || || ||
Db 102 TAACCTACCTTAATTAATTAATAAAAAAAAAAAAAA 66

```

RESULT 12

US-08-554-385-2/C
Sequence 2, Application US/08554385
Patent No. 6017692

GENERAL INFORMATION:

APPLICANT: Roger Brent
APPLICANT: Antonis S. Zervos
TITLE OF INVENTION: MAX-INTERACTING PROTEINS AND RELATED
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: Wordperfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/554,385
FILING DATE: No. 6017692ember 8, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:

NAME: Karen F. Lech
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/252001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 2417
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-554-385-2

```

Query Match          5.2%; Score 33; DB 3; Length 2417;
Best Local Similarity 58.8%; Pred. No. 1;
Matches 57; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

```

```

Qy 539 ggcacaccccttgggtccccccttgatgagcgctctgttcaactcattcaact 598
    || || || || || || || || || || || || || || || || || || ||
Db 162 GGAAGGCCCTTGGAGACCTTACCCTGGCTGGTGTGACTTTGTATTAATAATTT 103

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 07:21:10 ; Search time 443.18 Seconds
(without alignments)
2460.040 Million cell updates/sec

Title: US-09-319-156a-6
Perfect score: 635
Sequence: 1 cccgtatcttaactcctc.....tgaaaaaaaaaaaaaaa 635

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N.Geneseq_032802.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	635	100.0	635	19 AAV43215	Multiple sclerosis
2	635	100.0	635	20 AAV29702	Clone C16-3' from
3	617.4	97.2	2030	21 AAV63826	Nucleotide sequenc
4	545	85.8	1329	19 AAV43219	Multiple sclerosis
5	545	85.8	1329	20 AAV29704	Clone 5M6 from MSR
6	533.8	84.1	2946	20 AAV7526	Human secreted pro
7	533.8	84.1	2946	21 AAV59468	Human secreted pro
8	532.2	83.8	2781	22 AAV55650	Nucleotide sequenc
9	525.8	82.8	1136	20 AAV25660	Human endogenous r

10	525.8	82.8	1136	21 AAV59210
11	522.6	82.3	2782	20 AAV25661
12	522.6	82.3	2782	21 AAV59211
13	522.6	82.3	2782	22 AAV20069
14	511.4	80.5	7582	20 AAV25665
15	511.4	80.5	7582	21 AAV59215
16	499.6	78.7	1894	22 ABA45822
17	499.6	78.7	1894	22 ABA56337
18	499.6	78.7	1894	22 ABA25978
19	499.6	78.7	1894	22 ABA04516
20	499.6	78.7	1894	22 AAK30018
21	499.6	78.7	1894	22 AAI14608
22	499.6	78.7	1894	22 AAI35980
23	499.6	78.7	1894	22 AAI04422
24	495.6	78.0	3372	20 AAV25663
25	495.6	78.0	3372	21 AAV59213
26	493.2	77.7	2052	22 ABA08902
27	481.4	75.8	849	22 AAS31000
28	479.2	75.5	3831	23 AAS71727
29	479.2	75.5	5154	23 AAS67609
30	479.2	75.5	8279	23 AAS76474
31	479.2	75.5	8294	23 AAS84209
32	469.4	73.9	2942	23 AAS77313
33	467	73.5	7466	23 AAS68626
34	461.6	72.7	2527	22 AAI17481
35	460.8	72.6	583	23 AAS83931
36	453.4	71.4	1115	23 AAS76196
37	451	71.0	3903	22 AAI46211
38	449	70.7	1364	22 AAS63154
39	448.2	70.6	2300	23 AAS92680
40	446.2	70.3	1393	23 AAS31002
41	437.8	68.9	893	22 AAS65964
42	432.8	68.2	1666	23 AAS92500
43	422	66.5	1165	23 AAS72721
44	388.4	61.2	2385	22 AAI99411
45	388.4	61.2	2385	22 AAI63761

ALIGNMENTS

RESULT 1
AAV43215
ID AAV43215 standard; cDNA; 635 BP.
XX
AC AAV43215;
XX
DT 29-DEC-1998 (first entry)
XX
DE Multiple sclerosis associated retrovirus fragment 4.
XX
KM Multiple sclerosis associated retrovirus; MSRV; MS; pol gene;
KM 999 gene; env gene; rheumatoid arthritis-associated virus; 99.
XX
OS Multiple sclerosis associated retrovirus.
XX
FH Key Location/Qualifiers
FT CDS 1..234
FT /*tag= a
FT /product= "Encodes protein AAV71067"
XX
XX WO9823755-A1.
XX
XX 04-JUN-1998.
XX
XX 26-NOV-1997; 97WO-IB01482.
XX
XX 26-NOV-1996; 96US-0756429.
XX
XX (INMR) BIO MERIEUX.
XX
XX Bedin F, Beseme F, Jolivet-Reynaud C, Komurian-Pradel F;
PI Mandrand B, Paranhos-Baccala G, Perion H;

3' pol gene and 3'
Human endogenous r
5' non coding, 3'
HERV-W envelope pr
Complete human end
Human endogenous r
Human breast cell
Human foetal liver
Probe #4444 for ge
Human brain expres
Human bone marrow
Probe #4541 for ge
Probe #4666 used t
Probe #4413 used t
Human endogenous r
Partial pol gene a
Human receptor Tyr
Human diagnostic a
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
DNA encoding novel
Phosphoribosyl pyr
Human purified sec
DNA encoding novel
Human diagnostic a
DNA encoding novel
DNA encoding novel
DNA encoding novel
Human cDNA sequenc
DNA encoding novel
DNA encoding novel
Human excretory re
Human kidney relat

XX WPI: 1998-322732/28.
 DR P-PSDB: AAW71067.
 XX New nucleic acid from retroviruses - useful for diagnosis,
 PT prevention and treatment of, e.g. multiple sclerosis
 XX
 PS Disclosure: Page 183; 286pp; English.
 XX
 CC The present sequence represents a multiple sclerosis (MS) associated
 CC retrovirus (MSRV) genomic fragment used in the method of the
 CC invention. The invention provides complete or partial genomic
 CC sequences of the MSRV-1 pol gene, gag gene and env gene, and
 CC polyepitopes encoded by these genes. The invention also provides
 CC antibodies raised against the polyepitopes. The genomic sequences,
 CC polyepitopes and antibodies are also claimed useful for diagnosing
 CC infection by MS and rheumatoid arthritis-associated viruses, and also
 CC for prevention and treatment of infection with these viruses.
 XX
 Sequence 635 BP; 188 A; 170 C; 136 G; 141 T; 0 other;

Query Match 100.0%; Score 635; DB 19; Length 635;
 Best Local Similarity 100.0%; Pred. No. 2.2e-187;
 Matches 635; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cccgtatctttaaactcctctgttaagttgtctctccccaagaatcaaatgttaacta 60
 DB 1 cccgtatctttaaactcctctgttaagttgtctctccccaagaatcaaatgttaacta 60
 QY 61 caaatgttcttcaaatgagacacagatgagatccatgaactaagaccgtgagacc 120
 DB 61 caaatgttcttcaaatgagacacagatgagatccatgaactaagaccgtgagacc 120
 QY 121 ctgagacgagcctgtagcagacatgctcagatgttaatgacatgagagccctccgag 180
 DB 121 ctgagacgagcctgtagcagacatgctcagatgttaatgacatgagagccctccgag 180
 QY 181 gaaatctcaactgacaaacccctactatgcccgaatcagcgaggaagcagtgagcggt 240
 DB 181 gaaatctcaactgacaaacccctactatgcccgaatcagcgaggaagcagtgagcggt 240
 QY 241 catcagccaactcctcccaacagcactgtgttctctgttagagagggagctgagagac 300
 DB 241 catcagccaactcctcccaacagcactgtgttctctgttagagagggagctgagagac 300
 QY 301 aggaactagctgatttccctagcacaagaagaatccccaagcctgagcgagagtgact 360
 DB 301 aggaactagctgatttccctagcacaagaagaatccccaagcctgagcgagagtgact 360
 QY 361 gcatccacactcctaaacatgagggtctgcaactagctcacacccgacccaatcagaagagctc 420
 DB 361 gcatccacactcctaaacatgagggtctgcaactagctcacacccgacccaatcagaagagctc 420
 QY 421 actaaatgcttaattagcacaataaagagtgaaagaatagccaatcattatgctgctg 480
 DB 421 actaaatgcttaattagcacaataaagagtgaaagaatagccaatcattatgctgctg 480
 QY 481 agagacagcgaggaagcagaagatcgagatacaaacccaagcattcgagcggcagaagc 540
 DB 481 agagacagcgaggaagcagaagatcgagatacaaacccaagcattcgagcggcagaagc 540
 QY 541 caaaccccttgggtccctcccttctgtagcgctctgtttcactatcttcaactct 600
 DB 541 caaaccccttgggtccctccctcttctgtagcgctctgtttcactatcttcaactct 600
 QY 601 attaatcttgcaactgaaaaaaaaaaaaaaaaaaaaa 635
 DB 601 attaatcttgcaactgaaaaaaaaaaaaaaaaaaaaa 635

ID AAX29702 standard; DNA; 635 BP.
 XX
 AC AAX29702;
 XX
 DT 08-JUN-1999 (first entry)
 XX
 DE Clone CL6-3' from MSRV-1.
 XX
 KW Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
 XX rheumatoid polyarthritis; ss.
 OS Multiple sclerosis related virus type 1.
 XX
 PN FR2765588-A1.
 PD 08-JAN-1999.
 PE 07-JUL-1997; 97FR-0008816.
 PR 07-JUL-1997; 97FR-0008816.
 PA (INMR) BIO MERIEUX.
 DR WPI: 1999-098275/09.
 DR P-PSDB: AAW9552.
 XX
 PT Nucleic acid sequences of retrovirus called MSRV-1 - associated with
 PT multiple sclerosis or rheumatoid polyarthritis
 PS Claim 1: Page 36-37; 83pp; French.
 CC This sequence represents clone CL6-3' from a novel multiple sclerosis
 CC related virus type 1 (MSRV1). The sequence can be used in diagnostic,
 CC prophylactic or therapeutic compositions to inhibit expression of a
 CC multiple sclerosis related virus and/or virus associated with
 CC rheumatoid polyarthritis.
 XX
 SO Sequence 635 BP; 188 A; 170 C; 136 G; 141 T; 0 other;

QY 1 cccgtatctttaaactcctctgttaagttgtctctccccaagaatcaaatgttaacta 60
 DB 1 cccgtatctttaaactcctctgttaagttgtctctccccaagaatcaaatgttaacta 60
 QY 61 caaatgttcttcaaatgagacacagatgagatccatgaactaagaccgtgagacc 120
 DB 61 caaatgttcttcaaatgagacacagatgagatccatgaactaagaccgtgagacc 120
 QY 121 ctgagacgagcctgtagcagacatgctcagatgttaatgacatgagagccctccgag 180
 DB 121 ctgagacgagcctgtagcagacatgctcagatgttaatgacatgagagccctccgag 180
 QY 181 gaaatctcaactgacaaacccctactatgcccgaatcagcgaggaagcagtgagcggt 240
 DB 181 gaaatctcaactgacaaacccctactatgcccgaatcagcgaggaagcagtgagcggt 240
 QY 241 catcagccaactcctcccaacagcactgtgttctctgttagagagggagctgagagac 300
 DB 241 catcagccaactcctcccaacagcactgtgttctctgttagagagggagctgagagac 300
 QY 301 aggaactagctgatttccctagcacaagaagaatccccaagcctgagcgagagtgact 360
 DB 301 aggaactagctgatttccctagcacaagaagaatccccaagcctgagcgagagtgact 360
 QY 361 gcatccacactcctaaacatgagggtctgcaactagctcacacccgaaccaatcagaagagctc 420
 DB 361 gcatccacactcctaaacatgagggtctgcaactagctcacacccgaaccaatcagaagagctc 420
 QY 421 actaaatgcttaattagcacaataaagagtgaaagaatagccaatcattatgctgctg 480


```

|||||
Db 421 actaaatgctaattagcgaataaagagagtaagaataagccaatcatctatgctg 480
Qy 481 agagcaagcgaggagacaaagatcggtatataaaccaggaattcgagccggcaacg 540
Db 481 agagcaagcgaggagacaaagatcggtatataaaccaggaattcgagccggcaacg 540
Qy 541 caaccccttggtgccccccttgatgagcgctctgtttcactctattcactct 600
Db 541 caaccccttggtgccccccttgatgagcgctctgtttcactctattcactct 600
Qy 601 attaatctgcaactgcaaaaaaagagagagagagagagagagagagagagagag 635
Db 601 attaatctgcaactgcaaaaaaagagagagagagagagagagagagagagagag 635

RESULT 3
AAA63826
ID AAA63826 standard; DNA: 2030 BP.
XX AAA63826;
XX
DT 04-DEC-2000 (first entry)
XX
DE Nucleotide sequence of the MSRV-1 3' env and LTR regions.
XX
KM MSRV-1; pol region; long terminal repeat; LTR; RUS region; retrovirus;
KW ss.
XX
OS Multiple Sclerosis retrovirus 1.
XX
FH Key location/Qualifiers
FT CDS 1..1629
FT sig_peptide /tag= a
FT /note= "contains one termination codon"
FT /tag= b
FT CAAT_signal 1800..1807
FT /tag= c
FT CAAT_signal 1858..1864
FT /tag= d
FT TATA_signal 1906..1911
FT /tag= e
FT polyA_signal 1996..2002
FT /tag= f
XX
PN MO200047745-A1.
XX
D 17-AUG-2000.
XX
PE 15-FEB-2000; 2000MO-IB00159.
XX
PR 15-FEB-1999; 99EP-0420041.
XX
PA (INMR ) BIO MERIEUX.
XX
PI Paranhos-Baccala G, Perron H, Komurian-Pradel F;
XX
DR WPI: 2000-506097/45.
XX
DR P-PSDB; AAB08195.
XX
PT Nucleotide fragment of LTR-RUS region from Multiple Sclerosis
PT retrovirus (MSRV) used to detect the presence of MSRV-1 retrovirus in a
PT biological sample
XX
PS Disclosure; Fig 2; 23pp; English.
XX
CC The present sequence represents the nucleotide sequence corresponding
CC to the 3' env region and long terminal repeat sequences from clone
CC C16 of Multiple Sclerosis retrovirus (MSRV-1). The specification
CC describes a long terminal repeat (LTR)-RUS region which encodes the
CC expression of a MSRV-1 protein. This is unusual for LTRs, in
CC particular in the RUS region. The sequence includes CAAT and TATA

```

```

CC signals which are present in the U3 and R regions and are not directed
CC towards the CDS indicated in the features table. Probes and antibodies
CC to the MSRV-1 retrovirus protein and encoding polynucleotide sequences
CC are used to detect the presence of MSRV-1 retrovirus in a biological
CC sample.
XX
SQ Sequence 2030 BP; 574 A; 559 C; 387 G; 510 T; 0 other;

Query Match 97.2%; Score 617.4; DB 21; Length 2030;
Best Local Similarity 98.3%; Pred. No. 1..2e-181;
Matches 624; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 cccgtatcttcaacccttgtaagtgtgtcttccacgaatcaatgttaaaacta 60
Db 1396 cccgtatcttcaacccttgtaagtgtgtcttccacgaatcaatgttaaaacta 1455
Qy 61 caatgttctcaaatgagacacagatgagtcataagataacacgttgagacc 120
Db 1456 cagatgttctcaaatgagacacacagatgagtcataagataacacgttgagacc 1515
Qy 121 ctgagccgctctgtaagccatgctccgagttaattgaattgaagcacccctccgag 180
Db 1516 ctgagccgctctgtaagccatgctccgagttaattgaattgaagcacccctccgag 1575
Qy 181 gaatctcaactgacacacccctactatgcccacattcagcggaagcagttagagcgt 240
Db 1576 gaatctcaactgacacacccctactatgcccacattcagcggaagcagttagagcgt 1635
Qy 241 catcagccaactccccaacagcacttggtttctcgtttgagaggggagactgagagac 300
Db 1636 catcagccaactccccaacagcacttggtttctcgtttgagaggggagactgagagac 1695
Qy 301 aggactagctggaattctctgagccaagaaataccctaagctagctggagagtgact 360
Db 1696 aggactagctggaattctctgagccaagaaataccctaagctagctggagagtgact 1755
Qy 361 gcattccacttaaacatgaggcttgcaacttagctacaccccgacacacagagagctc 420
Db 1756 gcattccacttaaacatgaggcttgcaacttagctacaccccgacacacagagagctc 1815
Qy 421 actaaatgctaattagcgaataaagagagagtaagaagaatagccaatcatctatgctg 480
Db 1816 actaaatgctaattagcgaataaagagagagtaagaagaatagccaatcatctatgctg 1875
Qy 481 agagcaagcgaggagacaaagatcggtatataaaccaggaattcgagccggcaacg 540
Db 1876 agagcaagcgaggagacaaagatcggtatataaaccaggaattcgagccggcaacg 1935
Qy 541 caaccccttggtgccccccttgatgagcgctctgtttcactctattcactct 600
Db 1936 caaccccttggtgccccccttgatgagcgctctgtttcactctattcactct 1995
Qy 601 attaatctgcaactgcaaaaaaagagagagagagagagagagagagagagagag 635
Db 1996 attaatctgcaactgcaaaaaaagagagagagagagagagagagagagagagag 2030

RESULT 4
AAV43219
ID AAV43219 standard; cDNA: 1329 BP.
XX
AC AAV43219;
XX
DT 29-DEC-1998 (first entry)
XX
DE Multiple sclerosis associated retrovirus fragment 6.
XX
KM Multiple sclerosis associated retrovirus; MSRV; MS; pol gene;
KW gag gene; env gene; rheumatoid arthritis-associated virus; ss.
XX
OS Multiple sclerosis associated retrovirus.
XX

```

XX	Key	Location/Qualifiers
XX	CDS	2..490
FT		/*tag= a
FT		/product= "Encodes protein AAW71069"
FT		/transl_except= (pos:77-79, appears to code for a
FT		stop codon)
FT		/transl_except= (pos:125-127, appears to code for a
FT		stop codon)
FT		/transl_except= (pos:137-139, appears to code for a
FT		stop codon)
XX		
PN		MO9823755-A1.
PD		04-JUN-1998.
XX		
XX		26-NOV-1997; 97MO-IB01482.
XX		26-NOV-1996; 96US-0756429.
XX		
XX		(INMR) BIO MERIEUX.
PI		Bedin F, Beseme F, Jollivet-Reynaud C, Komurian-Pradel F,
PI		Mandrand B, Paranhos-Baccala G, Ferrion H;
XX		
DR		WPI; 1998-322732/28.
DR		P-PSDB; AAW71069.
XX		
XX		New nucleic acid from retroviruses - useful for diagnosis,
PT		prevention and treatment of, e.g. multiple sclerosis
XX		
XX		Disclosure; Pages 187-188; 286pp; English.
XX		
CC		The present sequence represents a multiple sclerosis (MS) associated
CC		retrovirus (MSRV) genomic fragment used in the method of the
CC		invention. The invention provides complete or partial genomic
CC		sequences of the MSRV-1 pol gene, gag gene and env gene, and
CC		polypeptides encoded by these genes. The invention also provides
CC		antibodies raised against the polypeptides. The genomic sequences,
CC		polypeptides and antibodies are also claimed useful for diagnosing
CC		infection by MS and rheumatoid arthritis-associated viruses, and also
CC		for prevention and treatment of infection with these viruses.
XX		
XX		Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;

Query Match	85.8%	Score 545	DB 19	length 1329
Best Local Similarity	92.4%	Pred. No.	3.3e-159	
Matches 572	Conservative	0	Mismatches 47	Indels 0
				Gaps 0

1 ccggtatctcttaaacccctctgttaattgtctcttcacgaatccaactgttaaaacta 60
Db 257 cccggtatctcttaaacctctctgtttaattgtctcttcacgaatctgaagctgttaagcta 316
Qy 61 caaatgtctcttcaaatcaggaacacagatctgaagtcacatgaataaccacgttgagacc 120
Db 317 caaatgtctcttcaaatcaggaacacacagatctgaagtcacatgaataactccgttgagacc 376
Qy 121 ctggacacggcctctctgaagccacatgctctccgaatgttaatgaacatctgaagacacccctcccgag 180
Db 377 ctggacacggcctctctgaagctatgtctctgaatgttaatgaacatctgaagtcacacccctcccgag 436
Qy 181 gaaatctcaactctgcacaacacccctactctatgcaccaatctcagctgaagacagcttgagacgct 240
Db 437 gaaatctcaactctgcacaacacccctactctacactccaattcagctgaagacagcttgagacagct 496
Qy 241 cactaagcgaacactccccaacacgcaactgggtttctccgttgtagagaggggacatcagagagac 300
Db 497 tgcctagccaacactccccaacacgtaactgggtttctccgttgtagagagggttgagacatcagagagac 556
Qy 301 aggaactagctgaatctctctgaagccaacagaaatcccttaagccatcagcttgagaaaggctgact 360
Db 557 aggaactagctgaatctctctgaagccatcagaaatccccaagcctatncttgagaaaggctgagacc 616

Oy	361	gcatccacccttaaacgttgggcttgtaacttaagctcaacccgacccaatcaagagctc	420
Db	617	gcatccactttaaacatagggtcttgtaacttagctcaacccgacccaatcaagagctc	676
Oy	421	actaaatgctaatataggcaaaaatagagagttaagaatagccaatcatctttgctg	480
Db	677	actaaatgctaatatcacagcaaaaacagagataagtaacatagccaatcatctttgctg	736
Oy	481	agaagacagcgaggagagacaagagatcgggatataaaccgcagcatcagacggacaag	540
Db	737	agagacacagcgaggagacaagagattggatataaactcagcatcuaagcagacaag	796
Oy	541	caacccctttgggtccctccctttgtataggcgctctgttttaactctatttcaact	600
Db	797	caacccctttgggtccctccctccatgtataggagctctgttttaactctatttcaactc	856
Oy	601	attaactctgcaactgaa	619
Db	857	attaactatgcaactgca	875

RESULT

ID AAX29704 standard; DNA; 1329 BP.

AC AAX29704 ;

DT 08-JUN-1999 (first entry)

Clone 5M6 from MSRV-1.

KW Multiple sclerosis; virus; di
rheumatoid polynephritis; se

Multiple sclerosis related virus type 1.

PN FR2765588-A1

PD 08-JAN-1999.

PF 07-JUL-1997; 97FR-0008816.

PR 07-JUL-1997; 97FR-0008816.

PA (INMR) BIO MERLEUX.
YY

DR WPL; 1999-098213/09
DR P-PSDB: AAW99554

PT Nucleic acid sequences of retrovirus called MSRV-1 - associated with

XXXXXXXXXXXXXXXXXXXXXXXXXXXX

XX

CC related virus type 1 (MSRV1). The sequence can be used in diagnostic,

CC multiple sclerosis related virus and/or virus associated with

[illegible]

SQ Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;

Query Match	85.88;	Score 545;	DB 20;	Length 1329;
-------------	--------	------------	--------	--------------

Matches 572; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 1 ccctgtatcttaacctctgttaagttgtctcttcagaatcaaaactgttaaacta 60

Db 257 ccctgtatcttcaacttcctgttaagttgtctctctccagaattgaagctgttaaagcta 316

Db 317 caaatgcttctcaaatggaaccccgagtcagtcacatgaactaaactacgctggagcc 376
 QY 121 ctgagccgctctgtagcccatgctcgcagatgtaatactagaagcaccctcccgag 180
 Db 377 ctgagccgctctgtagcccatgctcgcagatgtaatactagaagcaccctcccgag 436
 QY 181 gaaatctcaactgacacacccctactatgcccccaatcagcgggaagcagttagagcgt 240
 Db 437 gaatctcaactgacacacccctactactacactccaattcagtaggaagcagttagagcgt 436
 QY 241 catcagcaaacctcccaaaagcacttggcttctgttgaagaggagcagcagagac 300
 Db 497 tgcagcacaaccccccacaactgaggttctcgtgtgagagggtggagcagagac 556
 QY 301 agactagctggaattcctcagcgaacgaatccctcaagctagctgggaaggtgact 360
 Db 557 agactagctggaattcctcagcgaacgaatccctcaagctagctgggaaggtgact 616
 QY 361 gcatcacccttaacatactgggcttgcaacttagctacaccccgacccaacagagatc 420
 Db 617 gcatcacccttaacatactgggcttgcaacttagctacaccccgacccaacagagatc 676
 QY 421 actaaatgtctaattagcgaacaaatagaggagtaagaataagccaatcatctattgctg 480
 Db 677 actaaatgtctaattagcgaacaaatagaggagtaagaataagccaatcatctattgctg 736
 QY 481 agagcacagcggagaggaacagatcggaatataaacccagcagcttcgagccgcaacg 540
 Db 737 agagcacagcggagaggaacagatcggaatataaacccagcagcttcgagccgcaacg 796
 QY 541 caaccccttgggtccctccttctgtatcggcgctctgtttcactcatctactact 600
 Db 797 caaccccttgggtccctcctccttctgtatcggcgctctgtttcactcatctactact 856
 QY 601 attaatctgcaactgaa 619
 Db 857 attaatctgcaactgaa 875
 RESULT 6
 AAX77526
 ID AAX77526 standard; cDNA; 2946 BP.
 AC AAX77526;
 DT 10-AUG-1999 (first entry)
 DE Human secreted protein A172-2 cDNA.
 KW bone marrow; testes; brain; blood; placenta; human; murine; thymus;
 KW cell proliferation; prevention; nutrition; cytokine; immune; vaccine;
 KW hematopoiesis regulator; activin; inhibin; chemokine; chemokine;
 KW haemopoietic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;
 KW adherin; tumour invasion suppressor; gene therapy; tissue growth; ss.
 OS Homo sapiens.
 PN M09926972-A1.
 PD 03-JUN-1999.
 PE 17-NOV-1998; 98WO-US24614.
 PR 20-OCT-1998; 98US-0175928.
 PR 21-NOV-1997; 97US-0976110.
 PR 18-MAY-1998; 98US-0080478.
 PA (GEMV) GENETICS INST INC.
 PI Collins-Racie LA, Evans C, Jacobs K, Lavallie ER;
 PI McCoy JM, Merberg D, Treacy M;
 XX

DR MPI: 1999-357813/30.
 DR P-PSDB: AAY08622.
 XX
 PT New polynucleotides encoding secreted proteins
 XX
 PS Claim 13a; Page 100-101; 142pp; English.
 CC This invention describes novel human secreted proteins encoded by
 CC polynucleotides isolated from human adult testes, adult brain, adult
 CC blood or adult placenta, or murine adult bone marrow or thymus cDNA
 CC libraries. The products of the invention are predicted to have biological
 CC activities which would make them suitable for treating, preventing or
 CC ameliorating medical conditions in humans and animals, although no
 CC supporting data is given. Suggested activities include nutritional
 CC activity, cytokine and cell proliferation/differentiation activity,
 CC immune stimulating (e.g. as vaccines) or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemokine/chemokine activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, adherin/tumour invasion suppressor activity, and tumour
 CC inhibition activity. The polynucleotides are also stated to be useful
 CC for gene therapy.
 CC
 XX
 SQ Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other;
 Query Match 84.1%; Score 533.8; DB 20; Length 2946;
 Best Local Similarity 91.5%; Pred. NO. 1.5e-155;
 Matches 581; Conservative 0; Mismatches 42; Indels 12; Gaps 1;
 QY 1 cctgtatccttaacccctctgttaagttgtctctccagaatcaaaactgtaaacta 60
 Db 223 cctgtatccttaacccctctgttaagttgtctctccagaatcaaaactgtaaacta 2382
 QY 61 caaatgtcttcaaatgagacacagatgagtgcaatgactaacacgttgagcc 120
 Db 2383 -----caaatgagacacagatgagtgcaatgactaacacgttgagcc 2430
 QY 121 ctgagccgctctgtagcccatgctcgcagatgtaatactagaagcaccctcccgag 180
 Db 2431 ctgagccgctctgtagcccatgctcgcagatgtaatactagaagcaccctcccgag 2490
 QY 181 gaaatctcaactgacacacccctactatgcccccaatcagcgggaagcagttagcgt 240
 Db 2491 gaaatctcaactgacacacccctactatgcccccaatcagcgggaagcagttagcgt 2550
 QY 241 catcagcaaacctcccaaaagcacttggcttctgttgaagaggagcagagac 300
 Db 2551 cctcggcacaacccctcccaaaagcacttggcttctgttgaagaggagcagagac 2610
 QY 301 agactagctggaattcctcagcgaacgaatccctcaagctagctgggaaggtgact 360
 Db 2611 agactagctggaattcctcagcgaacgaatccctcaagctagctgggaaggtgact 2670
 QY 361 gcatcacccttaacatactgggcttgcaacttagctacaccccgacccaacagagatc 420
 Db 2671 gcatcacccttaacatactgggcttgcaacttagctacaccccgacccaacagagatc 2730
 QY 421 actaaatgtctaattagcgaacaaatagaggagtaagaataagccaatcatctattgctg 480
 Db 2731 actaaatgtctaattagcgaacaaatagaggagtaagaataagccaatcatctattgctg 2790
 QY 481 agagcacagcggagaggaacagatcggaatataaacccgagcttggagccggaacg 540
 Db 2791 agagcacagcggagaggaacagatcggaatataaacccgagcttggagccggaacg 2850
 QY 541 caaccccttgggtccctccttctgtatcggcgctctgtttcactcatctactact 600
 Db 2851 caaccccttgggtccctccttctgtatcggcgctctgtttcactcatctactact 2910
 QY 601 attaatctgcaactgaa 635
 Db 2911 attaatctgcaactgaa 2945

```
RESULT 7
AAZ59468
ID AAZ59468 standard; cDNA; 2946 BP.
XX
AC AAZ59468;
XX
DT 11-APR-2000 (first entry)
XX
DE Human secreted protein AJ172_2 polynucleotide sequence.
XX
KW Human; secreted protein; disease diagnosis; pre-eclampsia; cancer;
KW placental pathology; metastasis inhibition; nutritional activity;
KW immune stimulator; haematopoiesis regulator; tissue growth;
KW tumour inhibitor; anti-inflammatory; clone AJ172_2; ATCC_98115;
KW gene therapy; ss.
XX
OS Homo sapiens.
XX
MO9960020-A1.
XX
PD 25-NOV-1999.
XX
PF 17-MAY-1999; 99WO-US10915.
XX
PR 18-MAY-1998; 98US-0080478.
XX
PR 20-OCT-1998; 98US-0175928.
XX
PA (GENVY ) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, Laval11e ER, Collins-Racie LA, Evans C;
PI Mebergy D, M1 S, Treacy M;
XX
DR MPI: 2000-116311/10.
XX
P-PSDB: AAY67313.
XX
PT New polynucleotides encoding secreted cDNA libraries, used to develop
PS products for the diagnosis and treatment of neoplastic disease
XX
PS Claim 14; Page 107-108; 149pp; English.
XX
CC This is the human secreted protein AJ172_2 nucleotide sequence, obtained
CC from a human adult testes cDNA library. The invention relates to secreted
CC human and murine proteins. The polynucleotides and proteins are predicted
CC to have biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals. Detection of the levels of the proteins can be used for the
CC diagnosis of e.g. pre-eclampsia, placental pathology or cancer. Agents
CC which modulate the expression or function of the proteins may be used for
CC treating a neoplastic disease and inhibiting metastasis. Other suggested
CC activities include nutritional activity (e.g. in feeds), cytokine and
CC cell proliferation/differentiation activity, immune stimulating
CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
CC invasion suppressor activity, and tumour inhibition activity. The
CC polynucleotide sequences are also stated to be useful for gene therapy.
XX
SO Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other;

Query Match 84.1%; Score 533.8; DB 21; Length 2946;
Best Local Similarity 91.5%; Pred. No. 1.5e-155;
Matches 581; Conservative 0; Mismatches 42; Indels 12; Gaps 1;

OY 1 cccgtatcttaacctccctgttaagtgtctctcccaagaatcaaacgttaacta 60
DB 2333 cccgtatcttaacctccctgttaagtgtctctcccaagaatcaaacgttaacta 2382
OY 61 caaatgtcttcaatgagcaccagatgagtcacatgacatgaatgacaccgagacc 120
DB 61 caaatgtcttcaatgagcaccagatgagtcacatgacatgaatgacaccgagacc 120
```

```
DB 2383 -----caaatgagaccagaatgcagtcacagactaagaatctaccgagacc 2430
OY 121 ctgagccgctgtctagaccatgctccagatgttaatgacatggaagcaccctcccgag 180
DB 2431 ctgagccgctgtctagaccatgctctgtatgttaatgacatggaagcaccctcccgag 2490
OY 181 gaatctcaactgcacaacccctactatgcccacatcgaacggaagcagttagaagcgt 240
DB 2491 gaatctcaactgcacaacccctactatgcccacatcgaacggaagcagttagaagcgt 2550
OY 241 catagcccaactcccccacagacgtgtgtctctgttaaggggggagctgagagac 300
DB 2551 cgtcgccaacctcccccacagacgtgtgtctctgttaaggggggagctgagagac 2610
OY 301 aggaactagctgattctcctaagcacaagaatccctcaagcctagctggaaggtgact 360
DB 2611 aggaactagctgattctcctaagcacaagaatccctcaagcctagctggaaggtgact 2670
OY 361 gcatccaccttaaacatgaggctgcaacttagctcacaccgaccaatcagaagctc 420
DB 2671 acatccacctttaaacaagggctgcaacttagctcacaccgaccaatcagaagctc 2730
OY 421 actaaatgctaataggcaaaaatagaggttaagaatgccaatcattctatctcgt 480
DB 2731 actaaatgctaataggcaaaaatagaggttaagaatgccaatcattctatctcgt 2790
OY 481 agagcacagcgagaggaacagatcggagataaacccaagcattcggagcggaacgg 540
DB 2791 agagcacagcgagaggaacagatcggagataaacccaagcattcggagcggaacgg 2850
OY 541 caacccttgggtccctcccttgtatggcgctgtttcactctatcactc 600
DB 2851 caacccttgggtccctcccttgtatggcgctgtttcactctatcactc 2910
OY 601 attaatctgcaactgaaaaaataaaaaa 635
DB 2911 attaatctgcaactgaaaaaataaaaaa 2945

RESULT 8
AAF55630
ID AAF55630 standard; DNA; 2781 BP.
XX
AC AAF55630;
XX
DT 29-MAY-2001 (first entry)
XX
DE Nucleotide sequence of a human endogenous retrovirus envelope protein.
XX
KW Envelope protein; HERV; syncytia formation; placental development;
KW syncytia; cancer; cell adhesion; ss.
XX
OS Human endogenous retrovirus.
XX
FH key Location/Qualifiers
FH CDS 762..2378
FH FT /*tag= a
FH FT /product= "envelope protein"
XX
PN MO200116171-A1.
XX
PD 08-MAR-2001.
XX
PF 01-SEP-2000; 2000WO-FR02429.
XX
PR 01-SEP-1999; 99FR-0011141.
XX
PR 15-SEP-1999; 99FR-0011793.
XX
PA (INMR ) BIO MERIEUX.
XX
PA (INRM ) INST NAT SANTE & RECH MEDICALE.
XX
PI Mallet F, Cosset F, Blond J, Laval11ette D, Bouton O, Ruggieri A;
```



```

Db 798 aggaatctggtatcttcttaagctgactaagaatccctaaagctgagctgggaaggtgacc 857
Oy 361 gcatccacccctaaacaggggctgcaactagctcacccgacacatcagaagctc 420
Db 858 acatccacccctaaacaggggctgcaactagctcacccgacacatcagaagctc 917
Oy 421 actaaatgctaatatagcaaaaataggaagtaaagaatagccaatcatctatgctg 480
Db 918 actaaatgctaatatagcaaaaataggaagtaaagaatagccaatcatctatgctg 977
Oy 481 agagacagcgaggagggaggaatcgatcgatataaaccagacatcgagccggaacgg 540
Db 978 agagacagcgaggagggaggaatcgatcgatataaaccagacatcgagccggaacgg 1037
Oy 541 caaccctctgggctccctccctctgtagtgagcgctctgcttcaactctatcactc 600
Db 1038 caaccctctgggctccctccctctgtagtgagcgctctgcttcaactctatcactc 1097
Db 601 attaaatctgcaactgaaaaaataaaaaa 635
Db 1098 attaaatctgcaactgaaaaaataaaaaa 1132

```

RESULT 10

AAAS9210 standard; DNA; 1136 BP.

AAAS9210;

07-NOV-2000 (first entry)

3' pol gene and 3' non coding sequences of HERV-W from human genome.

Autoimmune disease: retrovirus; human endogenous retrovirus W; HERV-W; gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.

Homo sapiens.

MO200043521-A2.

27-JUL-2000.

21-JAN-2000; 2000WO-FR00144.

21-JAN-1999; 99FR-0000888.

(IMMR) BIO MERIEUX.

Paranhos-Baccala G, Mallet F, Volset C;

WPI: 2000-499229/44.

New nucleic acid from human endogenous retrovirus; useful e.g. for diagnosis of autoimmune disease and complications of pregnancy, contains at least part of the gag gene

Disclosure: Page 46; 53pp; French.

The present sequence represents an endogenous retroviral nucleic acid fragment, which is associated with an autoimmune disease, and is integrated into the human genome. The fragment is originally derived from a novel retrovirus, human endogenous retrovirus W (HERV-W). The HENV-W retrovirus is associated with autoimmune disease, failure of pregnancy or disorders of pregnancy. The nucleic acid fragment, or proteins derived from it, are useful for diagnosis of autoimmune disease (specifically multiple sclerosis) and for monitoring pregnancy. The nucleic acid fragments may also be used for in situ labelling of isolated chromosomes, while the transcription product can be used to study or monitor T cell proliferation in vitro.

Sequence 1136 BP; 336 A; 289 C; 241 G; 270 T; 0 other;

Query Match 82.8%; Score 525.8; DB 21; Length 1136;
Best Local Similarity 90.7%; Pred. No. 2.9e-153;
Matches 576; Conservative 0; Mismatches 47; Indels 12; Gaps 1;

```

Oy 1 cccgtatcttaacctctgttaagtgtctctcttccaaatcaaacgttaaata 60
Db 510 cccgtatcttaacctctgttaagtgtctctcttccaaatcaaacgttaaata 569
Oy 61 caaatgtcttcaaatgagcacagatgagatccatgaatcaccgagacc 120
Db 570 -----caaatgagaccaaagtgcagccaaatgaatgaatcaccgagacc 617
Oy 121 ctgagccgctctgtagcccatgctcgaagttaatgacatgagacccctccgag 180
Db 618 ctgagccgctctgtagcccatgctcgaagttaatgacatgagacccctccgag 677
Oy 181 gaatctcaatctgcaaacccctctatgcccacatcagggaggaagttaagcgt 240
Db 678 gaatctcaatctgcaaacccctctatgcccacatcagggaggaagttaagcgt 737
Oy 241 catcagccaacctccccaacagcactgtggttccgtgtgagaggggagctgagagac 300
Db 738 cgtcgccaacctccccaacagcactgtggttccgtgtgagagggagctgagagac 797
Oy 301 aggaatagctggtatctctagccaacgaaatccctaaagcctagctgggaagtgact 360
Db 798 aggaatagctggtatctctagccaacgaaatccctaaagcctagctgggaagtgact 857
Oy 361 gcatccacccctaaacaggggctgcaactagctcacccgacacatcagaagctc 420
Db 858 acatccacccctaaacaggggctgcaactagctcacccgacacatcagaagctc 917
Oy 421 actaaatgctaatatagcaaaaataggaagtaaagaatagccaatcatctatgctg 480
Db 918 actaaatgctaatatagcaaaaataggaagtaaagaatagccaatcatctatgctg 977
Oy 481 agagacagcgaggagggaggaatcgatcgatataaaccagacatcgagccggaacgg 540
Db 978 agagacagcgaggagggaggaatcgatcgatataaaccagacatcgagccggaacgg 1037
Oy 541 caaccctctgggctccctccctctgtagtgagcgctctgcttcaactctatcactc 600
Db 1038 caaccctctgggctccctccctctgtagtgagcgctctgcttcaactctatcactc 1097
Oy 601 attaaatctgcaactgaaaaaataaaaaa 635
Db 1098 attaaatctgcaactgaaaaaataaaaaa 1132

```

RESULT 11

AAAX25661 standard; cDNA to mRNA; 2782 BP.

AAAX25661;

21-MAY-1999 (first entry)

Human endogenous retrovirus W clone cl.PH74.

Clone: human endogenous retrovirus; genome: autoimmune disease; multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes; disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.

Human endogenous retrovirus.

W09902696-A1.

21-JAN-1999.

06-JUL-1998; 98WO-FR01442.

07-JUL-1997; 97FR-0000815.

XX (INMR) BIO MERIEUX.
 PA Beseme F, ~ Blond JL, Bouton O, Mallet F, Mandrand B;
 XX WPI: 1999-120897/10.
 XX
 DR New nucleic acid sequences from human endogenous retrovirus-W -
 XX expressed exclusively in placenta and useful in diagnosis and
 XX therapy of autoimmune disease, and abnormal or failed pregnancy
 XX
 PS Claim 1: Page 60-63; 106pp; French.
 CC This sequence represents clone cl.PH74 of the human endogenous retrovirus
 CC (HERV) W genome. The nucleic acids, their fragments or peptides encoded
 CC by them are markers of autoimmune disease (e.g. multiple sclerosis,
 CC rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-
 CC dependent diabetes and related pathologies) and of abnormal or
 CC unsuccessful pregnancy and can be used as chromosomal markers for
 CC susceptibility to these conditions, or proximity markers of genes
 CC associated with this susceptibility.
 SQ Sequence 2782 BP; 741 A; 767 C; 565 G; 709 T; 0 other;

Query Match 82.3%; Score 522.6; DB 20; Length 2782;
 Best Local Similarity 90.4%; Pred. No. 4,4e-152;
 Matches 574; Conservative 0; Mismatches 49; Indels 12; Gaps 1;

QY 1 ccctgtatctttaaactcctctgttaagtgtctctccagaatcaaaactgtanaacta 60
 DB 2158 ccctgtatctttaaactcctctgttaagtgtctctccagaatcaaaactgtanaacta 2217
 QY 61 caattgtctttaaactcctctgttaagtgtctctccagaatcaaaactgtanaacta 120
 DB 2218 -----caaatggagcccaagatgacgtccaaactgaactccagagacc 2265
 QY 121 ctggaccggtctgtacatgctccgatgttaactgtacactgaaggcaccctccagag 180
 DB 2266 ctggaccggtctgtacatgctccgatgttaactgtacactgaaggcaccctccagag 2325
 QY 181 gaaatcctaactgcacaacccctactatgcccgaatcagcgggaagcagttagagcgt 240
 DB 2326 gaaatcctaactgcacaacccctactatgcccgaatcagcgggaagcagttagagcgt 2385
 QY 241 catcagcacaacctcccaagaagcactgtgttctcgtgtgaagaggggagcagagac 300
 DB 2386 ggtcggccaaacctcccaagaagcactgtgttctcgtgtgaagaggggagcagagac 2445
 QY 301 aggaactagctgaatttctcctagcacaagaatccctaagctagctgggaaggtgac 360
 DB 2446 aggaactagctgaatttctcctagcacaagaatccctaagctagctgggaaggtgac 2505
 QY 361 gcatccaacttaacaatgggctgtgcaactagctcacaccgacccaatcagagagctc 420
 DB 2506 acatccaactttaaaccgggctgtgcaactagctcacaccgacccaatcagagagctc 2565
 QY 421 actaaatgttaattggcaaaataggagtaagaagaatagccaatcattatgctgt 480
 DB 2566 actaaatgttaattggcaaaataggagtaagaagaatagccaatcattatgctgt 2625
 QY 481 agagcacaagcggagagacagatcggtatataaccccgactgtgagccggcagcgg 540
 DB 2626 agagcacaagcggagagacatgctggatataaccccgactgtgagccggcagcgg 2665
 QY 541 caaccctcttggtccctcccttctatggcgctcgtcttctactctattactact 600
 DB 2686 caaccctcttggtccctcccttctatggcgctcgtcttctactctattactact 2745
 QY 601 attaatctgtcaactgtaaaaaataaaaaa 635
 DB 2746 attaatctgtcaactgtcaaaaaaataaaaaa 2780

RESULT 12
 ID AAA59211 standard; DNA; 2782 BP.
 XX
 AC AAA59211;
 XX
 DT 07-NOV-2000 (first entry)
 XX
 DE 5' non coding, 3' pol gene, and 3' non coding sequences of HERV-W.
 XX
 KW Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
 KM gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200043521-A2.
 XX
 PD 27-JUL-2000.
 XX
 PF 21-JAN-2000; 2000MO-FR00144.
 XX
 PR 21-JAN-1999; 99FR-0000888.
 XX
 PA (INMR) BIO MERIEUX.
 XX
 PI Paranhos-Baccala G, Mallet F, Voisset C;
 XX
 DR WPI: 2000-499229/44.
 XX

New nucleic acid from human endogenous retrovirus, useful e.g. for
 diagnosis of autoimmune disease and complications of pregnancy,
 contains at least part of the gag gene -
 Disclosure: Page 46-47; 53pp; French.

The present sequence represents an endogenous retroviral nucleic acid
 fragment, which is associated with an autoimmune disease, and is
 integrated into the human genome. The fragment is originally derived
 from a novel retrovirus, human endogenous retrovirus W (HERV-W). The
 HERV-W retrovirus is associated with autoimmune disease, failure of
 pregnancy or disorders of pregnancy. The nucleic acid fragment, or
 proteins derived from it, are useful for diagnosis of autoimmune
 disease (specifically multiple sclerosis) and for monitoring pregnancy.
 The nucleic acid fragments may also be used for in situ labelling of
 isolated chromosomes, while the transcription product can be used to
 study or monitor T cell proliferation in vitro.

Sequence 2782 BP; 741 A; 767 C; 565 G; 709 T; 0 other;

Query Match 82.3%; Score 522.6; DB 21; Length 2782;
 Best Local Similarity 90.4%; Pred. No. 4,4e-152;
 Matches 574; Conservative 0; Mismatches 49; Indels 12; Gaps 1;

QY 1 ccctgtatctttaaactcctctgttaagtgtctctccagaatcaaaactgtanaacta 60
 DB 2158 ccctgtatctttaaactcctctgttaagtgtctctccagaatcaaaactgtanaacta 2217
 QY 61 caattgtctttaaactcctctgttaagtgtctctccagaatcaaaactgtanaacta 120
 DB 2218 -----caaatggagcccaagatgacgtccaaactgaacttaacgagacc 2265
 QY 121 ctggaccggtctgtacatgctccgatgttaactgtacactgaaggcaccctccagag 180
 DB 2266 ctggaccggtctgtacatgctccgatgttaactgtacactgaaggcaccctccagag 2325
 QY 181 gaaatcctaactgcacaacccctactatgcccgaatcagcgggaagcagttagagcgt 240
 DB 2326 gaaatcctaactgcacaacccctactatgcccgaatcagcgggaagcagttagagcgt 2385
 QY 241 catcagcacaacctcccaagaagcactgtgttctcgtgtgaagaggggagcagagac 300
 DB 2445 catcagcacaacctcccaagaagcactgtgttctcgtgtgaagaggggagcagagac 2780

QY 421 actaaatgcttaattagcgcaaaaataagagtaaaagaatagccaatcatctatgtcgtg 480
 |||||||
 Db 2566 actaaatgcttaattagcgcaaaaataagagtaaaagaatagccaatcatctatgtcgtg 2675
 |||||||
 QY 481 agagcacaagcgaggaagacaaagatcgagatataaaccagcaattcgagccgcaacg 540
 |||||||
 Db 2626 agagcacaagcgaggaagacaaagatcgagatataaaccagcaattcgagccgcaacg 2685
 |||||||
 QY 541 caaccccttgggtccctcccttggatgagcgctctgttttcaatctattcaactc 600
 |||||||
 Db 2686 caaccccttgggtccctcccttggatgagcgctctgttttcaatctattcaactc 2745
 |||||||
 QY 601 attaatcttgcacactgaataaaaaa 635
 |||||||
 Db 2746 attaatcttgcacactgaataaaaaa 2780
 |||||||
 RESULT 14
 AAX25665
 AAX25665 standard: cDNA to mRNA; 7582 BP.
 AAX25665:
 AC AAX25665:
 DT 21-MAY-1999 (first entry)
 DE Complete human endogenous retrovirus W genome.
 KW Clone: human endogenous retrovirus; genome; autoimmune disease;
 KM multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
 KW disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
 OS Human endogenous retrovirus.
 XX WO902696-A1.
 PN 21-JAN-1999.
 PD 21-JAN-1999.
 XX 06-JUL-1998: 98MO-FR01442.
 PF 07-JUL-1997: 97FR-0008815.
 PR (INMR) BIO MERIEUX.
 PA Beseme F, Blond JL, Bouton O, Mallet F, Mandrand B;
 PI WPI: 1999-120897/10.
 DR WPI: 1999-120897/10.
 XX New nucleic acid sequences from human endogenous retrovirus-W -
 expressed exclusively in placenta and useful in diagnosis and
 therapy of autoimmune disease, and abnormal or failed pregnancy
 Claim 1; Page 71-74; 106pp; French.
 PS This sequence represents the complete sequence of the human endogenous
 CC retrovirus (HERV) W genome. The nucleic acids, their fragments or
 CC peptides encoded by them are markers of autoimmune disease (e.g. multiple
 CC sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus,
 CC insulin-dependent diabetes and related pathologies) and of abnormal or
 CC unsuccessful pregnancy and can be used as chromosomal markers for
 CC susceptibility to these conditions, or proximity markers of genes
 CC associated with this susceptibility.
 XX
 SO Sequence 7582 BP; 2156 A; 1877 C; 1537 G; 1796 T; 2 U; 214 other;

QY 61 caaatgtcttccaatggagcaaccagatgagtcacatgaactaagatccacgtgagccc 120
 |||||||
 Db 7036 -----caaatggagcccaagatgagtcacatgaactaagatccacgtgagccc 7083
 |||||||
 QY 121 ctgagccggtcgtctagaccatgtctccagatgttaatgacatgaaagcaccctccag 180
 |||||||
 Db 7084 ctgagccggtcgtctagaccatgtctccagatgttaatgacatgaaagcaccctccag 7143
 |||||||
 QY 181 gaaatctcaactgacaaaccctactatgccccaatctcagcgggaagcagttgagcggt 240
 |||||||
 Db 7144 gaaatctcaactgacaaaccctactatgccccaatctcagcgggaagcagttgagcggt 7203
 |||||||
 QY 241 catcagcacaaccccccacaagacattgggttttccgtgtgagaggggagcagagagac 300
 : || |||||
 Db 7204 gtcgagcacaaccccccacaagacattgggttttccgtgtgagaggggagcagagagac 7263
 |||||||
 QY 301 aggaactagctgatttccctagcacaagaaatccctaaagcctagctggagagtgact 360
 |||||||
 Db 7264 aggaactagctgatttccctagcacaagaaatccctaaagcctagctggagagtgact 7323
 |||||||
 QY 361 gcatcaccctctaacaactgggtctgcaacttagctcacaccgacccaatcagagagctc 420
 |||||||
 Db 7324 acatccaccctctaacaacggtgctgcaacttagctcacaccgacccaatcagagagctc 7383
 |||||||
 QY 421 actaaatgcttaattagcgcaaaaataagagtaaaagaatagccaatcatctatgtcgtg 480
 |||||||
 Db 7384 actaaatgcttaattagcgcgcaaaaataagagtaaaagaatagccaatcatctatgtcgmty 7443
 |||||||
 QY 481 agagcacaagcgaggaagacaaagatcgagatataaaccagcaattcgagccgcaacg 540
 |||||||
 Db 7444 agagcacaagcgaggaagacaaagatcgagatataaaccagcaattcgagccgcaacg 7503
 |||||||
 QY 541 caaccccttgggtccctcccttggatgagcgctctgttttcaatctattcaactc 600
 |||||||
 Db 7504 caaccccttgggtccctcccttggatgagcgctctgttttcaatctattcaactc 7563
 |||||||
 QY 601 attaatcttgcacactg 617
 |||||||
 Db 7564 attaatcttgcacactg 7580
 |||||||
 RESULT 15
 AAA59215
 ID AAA59215 standard: DNA; 7582 BP.
 XX AAA59215:
 AC AAA59215:
 DT 07-NOV-2000 (first entry)
 DE Human endogenous retrovirus W (HERV-W) sequence.
 KW Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
 KM gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
 OS Human endogenous retrovirus.
 XX
 FH Key Location/Qualifiers
 FT LTR 1..120
 FT /*tag- a
 FT /*tag- "R of 5' LTR"
 FT 121..575
 FT /*tag- b
 FT /*note- "U5 of 5' LTR"
 FT 579..596
 FT primer_bind
 FT /*tag- c
 FT 5581..7194
 FT /*tag- d
 FT /*note- "ORF1 env538"
 FT 7039..7194
 FT /*tag- e
 FT /*note- "ORF2 52 AA"
 FT 7112..7255
 FT CDS

```

FT      /tag= f
FT      /note= "ORF3 48 AA"
FT      misc-feature
FT      7244..7254
FT      /tag= g
FT      /note= "polyurine tract"
FT      LTR
FT      7256..7582
FT      /tag= h
FT      /note= "U3-R of 3' LTR"
FT      polyA_signal
FT      7563..7569
FT      /tag= 1
XX      WO20043521-A2.
XX      27-JUL-2000.
XX      21-JAN-2000; 2000WO-FR00144.
XX      21-JAN-1999; 99FR-0000888.
XX      (INMR ) BIO MERIEUX.
XX      Paranhos-Bacala G, Mallet F, Voliset C;
XX      WPI: 2000-499229/44.
XX      New nucleic acid from human endogenous retrovirus, useful e.g. for
XX      diagnosis of autoimmune disease and complications of pregnancy,
XX      contains at least part of the gag gene
XX      PS      Disclosure: Page 49-52; 53pp; French.
XX      CC      The present sequence represents an endogenous retrovirus, which is
XX      associated with an autoimmune disease, and is integrated into the human
XX      genome. The retrovirus is human endogenous retrovirus W (HERV-W). The
XX      HERV-W retrovirus is associated with autoimmune disease, failure of
XX      pregnancy or disorders of pregnancy. HERV-W nucleic acid fragments, or
XX      proteins derived from it, are useful for diagnosis of autoimmune
XX      disease (specifically multiple sclerosis) and for monitoring pregnancy.
XX      The nucleic acid fragments may also be used for in situ labelling of
XX      isolated chromosomes, while the transcription product can be used to
XX      study or monitor T cell proliferation in vitro.
XX      SO      Sequence 7582 BP: 2156 A; 1876 C; 1538 G; 1796 T; 216 other;

Query Match      80.5%; Score 511.4; DB 21; Length 7582;
Best Local Similarity 89.5%; Pred. No. 2,2e-148;
Matches 552; Conservative 11; Mismatches 42; Indels 12; Gaps 1;

1 cccgtatcttaacctcctgttaagttgtctcttccagaatcaaaactgtaacta 60
Db      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
6976 cccgtatctttracctctgttaacttctcttccagaatcgaagcttiraacta 7035
OY      61 caaatgtcttcaaatgagcagatgagatcagatcaatcaccgtgagacc 120
Db      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
7036 -----caaatgagcccaagatgcagtcacaaagactaactaccgcagacc 7083
OY      121 ctgagccgctctagcccatgctccatgttaatgacatgaaagcaccctccag 180
Db      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
7084 ctgagccgctctgagcccatgctgttaatgacatcaaaagcaccctctctgag 7143
OY      181 gaaatcctaactgcaaacctctactatgcccnaatlcagcgggaagcagttagaagcgt 240
Db      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
7144 gaaatcctaactgcaaacctctactatgcccnaatlcagcgggaagcagttagaagcgt 7203
OY      241 catcagccaacccctcagcagacttggtttctctgttgaaggaggagctgagaagc 300
Db      :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
7204 gctcggccaacccctcagcagacttggtttctctgttgaaggaggagctgagaagc 7263
OY      301 aggaactagctggaattctcttaggccaagaaatcccttaagccttagctggaagtgact 360
Db      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
7264 aggaactagctggaattctcttaggccaagaaatcccttaagccttagctggaagtgact 7323

```

```

OY      361 gcatccaccttaacatggtggtcgaacttagctcacaccgacccaatcagagagctc 420
Db      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
7324 acatccacctttaaacaacggtcttgcaactagylcaacacctgacccaatcagagagctc 7383
OY      421 actaaatgcttaattagggcaaaaatagagggaagaagaatagccaatcattatggcctg 480
Db      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
7384 actaaatgcttaattagggcaaaaatagagggaagaagaatagccaatcattatggcctg 7443
OY      481 agagcacagcggaggaagacaagagatcgagataaaaccagcatcgcagccgcaacg 540
Db      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
7444 agagcacagcggaggaagacaagagatcgagatataaacccaagtcttcgagccgcaacg 7503
OY      541 caacccttggtggtccctcccttgtatggtggtcgtgttcttcaactatcactct 600
Db      |||||||:|||||:|||||:|||||:|||||:|||||:|||||:
7504 caacccttggtggtccctcccttgtatggtggtcgtgttcttcaactatcactct 7563
OY      601 attaacttgcaactg 617
Db      |||||||:|||||:|||||:|||||:|||||:|||||:
7564 attaacttgcaactg 7580

```

Search completed: June 20, 2002, 09:42:19
Job time: 8469 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 06:44:35 ; Search time 4429.18 Seconds
(without alignments)
3000.183 Million cell updates/sec

Title: US-09-319-156a-6

Perfect score: 635
Sequence: 1 ccctgtatcttcaactctct.....tgaaaaaaaaaaaaaaaaa 635

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_da: *
2: gb_da: *
3: gb_in: *
4: gb_in: *
5: gb_ov: *
6: gb_ov: *
7: gb_ph: *
8: gb_ph: *
9: gb_pl: *
10: gb_pl: *
11: gb_to: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_mu: *
21: em_or: *
22: em_ov: *
23: em_ov: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vi: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_inv: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result Query Match Length DB ID Description

1	635	100.0	635	6	AX001024	AX001024 Sequence
2	635	100.0	2004	14	AF127229	AF127229 Multiple
3	631.8	99.5	176095	2	AC019346	AC019346 Homo sapi
4	631.8	99.5	210336	2	AC022171	AC022171 Homo sapi
5	580.2	91.4	148724	2	AC010778	AC010778 Homo sapi
6	579.2	91.2	8339	2	AL162912	AL162912 Human DNA
7	571	89.9	136901	9	AC073626	AC073626 Homo sapi
8	564.6	88.9	89728	9	AL583805	AL583805 Human DNA
9	564.6	88.9	212096	2	AL451130	AL451130 Homo sapi
10	560.2	88.2	175040	2	AC072023	AC072023 Homo sapi
11	556.6	87.7	174019	2	AP001538	AP001538 Homo sapi
12	556.6	87.7	340000	2	AP001674	AP001674 Homo sapi
13	551.8	86.9	149755	2	AP001545	AP001545 Homo sapi
14	551.8	86.9	152775	2	AC090313	AC090313 Homo sapi
15	551.8	86.9	163803	2	AC093531	AC093531 Homo sapi
16	551.8	86.9	179109	2	AC107075	AC107075 Homo sapi
17	551.4	86.8	128468	9	AL135090	AL135090 Human DNA
18	551.4	86.8	140756	2	AL135038	AL135038 Human DNA
19	551.4	86.8	167366	2	AC021774	AC021774 Homo sapi
20	551.4	86.8	183499	2	AL607153	AL607153 Homo sapi
21	550.2	86.6	164211	2	AP002790	AP002790 Homo sapi
22	550.2	86.6	168055	2	AC064801	AC064801 Homo sapi
23	549	86.5	180523	2	CNS01DSH	AL121784 Human chr
24	549	86.5	190565	2	AC007374	AC007374 Homo sapi
25	547.6	86.2	192178	2	AC009443	AC009443 Homo sapi
26	547.6	86.2	192899	2	CNS05TF5	AL365295 Human chr
27	547.4	86.2	172281	9	AC068492	AC068492 Homo sapi
28	547.2	86.2	168195	2	AC104163	AC104163 Homo sapi
29	546.6	86.1	158564	2	AC104444	AC104444 Homo sapi
30	546	86.0	169234	2	CNS06C7Q	AL139079 Human chr
31	545.6	85.9	170730	30	AC073868	AC073868 Homo sapi
32	545	85.8	1329	6	AX001030	AX001030 Sequence
33	545	85.8	99408	2	AL359385	AL359385 Human DNA
34	545	85.8	170586	2	AL356126	AL356126 Homo sapi
35	543.8	85.6	197043	2	AC087897	AC087897 Homo sapi
36	543.4	85.6	88328	2	AL357874	AL357874 Human DNA
37	538.8	84.9	122969	2	AC105029	AC105029 Homo sapi
38	537.8	84.7	83412	9	AC092843	AC092843 Homo sapi
39	537.8	84.7	176600	9	AC046130	AC046130 Homo sapi
40	534.4	84.2	175923	2	AC064839	AC064839 Homo sapi
41	534.4	84.2	187289	2	AC083908	AC083908 Homo sapi
42	533.8	84.1	2946	6	ARI77269	ARI77269 Sequence
43	532.2	83.8	2781	6	AX092223	AX092223 Sequence
44	532.2	83.8	2781	6	AF072506	AF072506 Homo sapi
45	532	83.8	157711	9	AC055738	AC055738 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS AX001024 635 bp DNA linear PAT 10-MAR-2000
DEFINITION Sequence 6 from Patent WO902666.
ACCESSION AX001024
VERSION AX001024.1 GI:7241262

SOURCE
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 635)
AUTHORS Otc.C. and Bedin.F.
TITLE RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN PARTICULAR
ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR RHEUMATOID ARTHRITIS, FOR
DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES
JOURNAL Patent: WO 9902666-A 6 21-JAN-1999;
BIO MERIEUX (FR); OTT CATHERINE
FEATURES
1..635 Location/Qualifiers
BASE COUNT 188 a 170 c 136 g 141 t

ORIGIN
/organism="unidentified"
/db_xref="taxon:32644"

Query Match 100.0%; Score 635; DB 6; Length 635;
Best Local Similarity 100.0%; Pred. No. 3,3e-192;
Matches 635; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 cccgtatcttcaacctctgttaagttgtctctccagaatcaaaactgtaaaacta 60
    |||||||
DB 1 CCTGTATCTTTAACTCCTTGTAAAGTTGCTCTCCAGATCAAAACTGTAAACTA 60
OY 61 caaatgtcttcaaaatgagacacagatgagatcagatcaagatccacgtgagacc 120
    |||||||
DB 61 CAATATTGTTCTTCAAAATGAGACACAGATGATCCATCACTAAGATCCACGTGAGACC 120
OY 121 ctgagacggcctgtagccatgctccga tgttaatgacatgaaagcaacctcccgag 180
    |||||||
DB 121 CTGGACCGGCGCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGACACCCCTCCGAG 180
OY 181 gaattcaactgacacacacctactatgccccaatcagcggaagcaagtagagcggt 240
    |||||||
DB 181 GAAATCTCAACTGACACACCCCTACTATGCCCAATTTCAGCGGGAAGCAGTTAGACCGGT 240
OY 241 catcagcaaacctcccaacagacacttgggtttctctgttgaaggggagactgaaagac 300
    |||||||
DB 241 CATCAGCCAACTCCCCACAGACACTTGGGTTTCTGTTGAGAGGGGAGCTGAGAGAC 300
OY 301 aggaactagctggaattctcctagggccaagaaatccctaaagcttgggaagtgact 360
    |||||||
DB 301 AGGACTAGCTGGAATTCTTAGGCCAAGAAATCCCTTAAGCTACTGGAAGGTGACT 360
OY 361 gcatcacccttcaaaatgaggttgcaactagctcaacccgacaaatcagaagagctc 420
    |||||||
DB 361 GCATCCACCTCTAAMATGAGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 420
OY 421 actaaatgctaaataggaataagaggtcaagaataagccaatcactatctgctg 480
    |||||||
DB 421 ACTAAATGCTTAATTAGGCAAAAATAGAGGTAAGAAATAGCCATCTCTATTGCTCTG 480
OY 481 agagacacagcgaggaagacaaagatcggagataaaacccagacatcgaacggcaacgg 540
    |||||||
DB 481 AGAGCACACGGGAGGAGACAGATCGGATATTAAMCCAGGACATTCGAGCGGCAACGG 540
OY 541 caaccctcttggttccctcccttctgtatggcgctcgttcttcaactatctcaact 600
    |||||||
DB 541 CAACCCCTTTGGTGCCCTCCCTTTGTATGGCGCTCTGTATTCTACTCTCT 600
OY 601 attaaatctgcaactgtaaaataaaataaaataaa 635
    |||||||
DB 601 ATTAAATCTTGCAACTGAAAAAATAAAAAA 635
```

RESULT 2
AF127229 2004 bp mRNA linear VRL 11-AUG-1999

LOCUS AF127229 Multiple sclerosis associated retrovirus element clone Cl6
DEFINITION pol-emb/3'LTR-like mRNA sequence.

ACCESSION AF127229
VERSION AF127229.1 GI:5726294

KEYWORDS multiple sclerosis associated retrovirus element.
SOURCE multiple sclerosis associated retrovirus element.
ORGANISM Viruses; Retroid viruses; Retroviridae.

REFERENCE 1 (bases 1 to 2004)
AUTHORS Ounanian-Paraz,A., Sodoyer,M., Ott,C., Rajoharison,A., Garcia,E.,

Mallet,F., Mandrand,B. and Perron,H.

TITLE Molecular cloning and characterization of MSRV-related sequences

JOURNAL Virology 260 (1), 1-9 (1999)

PUBMED 10405350

REFERENCE 2 (bases 1 to 2004)
AUTHORS Komurian-Pradel,F., Paranhos-Bacalca,G., Bedin,F.,

Ounanian-Paraz,A., Sodoyer,M., Ott,C., Rajoharison,A., Garcia,E.,
Mallet,F., Mandrand,B. and Perron,H.
Direct Submission
Submitted (10-FEB-1999) UMR103 CNRS, bioMerieux, 46, allée
D'italie, Lyon 69007, France

FEATURES
source location/Qualifiers
1..2004
/organism="multiple sclerosis associated retrovirus
element"
/db_xref="taxon:89382"
/clone="Cl6"

misc_feature 1..2004
/note="Similar to pol-emb and 3'LTR region"

BASE COUNT 641 a 476 c 438 g 449 t

ORIGIN

Query Match 100.0%; Score 635; DB 14; Length 2004;
Best Local Similarity 100.0%; Pred. No. 4,2e-192;
Matches 635; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 cccgtatcttcaacctctgttaagttgtctctccagaatcaaaactgtaaaacta 60
    |||||||
DB 1370 CCTGTATCTTTAACTCCTTGTAAAGTTGCTCTCCAGATCAAAACTGTAAACTA 1429
OY 61 caaatgtcttcaaaatgagacacagatgagatcagatcaagatccacgtgagacc 120
    |||||||
DB 1430 CAATATTGTTCTTCAAAATGAGACACAGATGATCCATCACTAAGATCCACGTGAGACC 1489
OY 121 ctgagacggcctgtagccatgctccga tgttaatgacatgaaagcaacctcccgag 180
    |||||||
DB 1490 CTGGACCGGCGCTGCTAGCCCATGCTCCGATGTTAATGACATTGAAGACACCCCTCCGAG 1549
OY 181 gaattcaactgacacacacctactatgccccaatcagcggaagcaagtagagcggt 240
    |||||||
DB 1550 GAAATCTCAACTGACACACCCCTACTATGCCCAATTTCAGCGGGAAGCAGTTAGACCGGT 1609
OY 241 catcagcaaacctcccaacagacacttgggtttctctgttgaaggggagactgaaagac 300
    |||||||
DB 1610 CATCAGCCAACTCCCCACAGACACTTGGGTTTCTGTTGAGAGGGGAGCTGAGAGAC 1669
OY 301 aggaactagctggaattctcctagggccaagaaatccctaaagcttgggaagtgact 360
    |||||||
DB 1670 AGGACTAGCTGGAATTCTTAGGCCAAGAAATCCCTTAAGCTACTGGAAGGTGACT 1729
OY 361 gcatcacccttcaaaatgaggttgcaactagctcaacccgacaaatcagaagagctc 420
    |||||||
DB 1730 GCATCCACCTCTAAMATGAGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 1789
OY 421 actaaatgctaaataggaataagaggtcaagaataagccaatcactatctgctg 480
    |||||||
DB 1790 ACTAAATGCTTAATTAGGCAAAAATAGAGGTAAGAAATAGCCATCTCTATTGCTCTG 1849
OY 481 agagacacagcgaggaagacaaagatcggagataaaacccagacatcgaacggcaacgg 540
    |||||||
DB 1850 AGAGCACACGGGAGGAGACAGATCGGATATTAAMCCAGGACATTCGAGCGGCAACGG 1909
OY 541 caaccctcttggttccctcccttctgtatggcgctcgttcttcaactatctcaact 600
    |||||||
DB 1910 CAACCCCTTTGGTGCCCTCCCTTTGTATGGCGCTCTGTATTCTACTCTCT 1969
OY 601 attaaatctgcaactgtaaaataaaataaaataaa 635
    |||||||
DB 1970 ATTAAATCTTGCAACTGAAAAAATAAAAAA 2004
```

RESULT 3
AC019346 176095 bp DNA linear HTG 25-OCT-2001

LOCUS AC019346/C Homo sapiens chromosome 18 clone RP11-497M7 map 18, WORKING DRAFT

DEFINITION SEQUENCE, 1 ordered pieces.

ACCESSION AC019346
VERSION AC019346.4 GI:16418201

KEYWORDS HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 18, clone RP11-497M7
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 176095)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F., Boguslavsky, L., Boukhaltier, B., Brown, A., Burkett, G., Castle, A., Chepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Deaellano, K., Dewar, K., Domino, M., Doyle, M., Fenesor, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lehocsky, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Pierre, N., Pisanl, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tefaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (02-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Oct 25, 2001 this sequence version replaced g1:7230183. All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Center project name: L4296
Center clone name: 497_M_7

Summary Statistics
Sequencing vector: M13; M7815; 32% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 176087 bases at least Q40
Consensus quality: 176095 bases at least Q30
Insert size: 178000; agarose-fp
Insert size: 176095; sum-of-ctrls
Quality coverage: 11.5 in Q20 bases; agarose-fp
Quality coverage: 11.7 in Q20.

NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 176095: contig of 176095 bp in length.
Assembly-fragment.

Location/Qualifiers
1. 176095
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-497M7"
/clone_lib="RP11 Human Male BAC"

BASE COUNT 56205 a 30885 c 31099 g 57906 t
ORIGIN

Query Match 99.5%; Score 631.8; DB 2; Length 176095;
Best Local Similarity 99.7%; Pred. No. 1.1e-190;
Matches 633; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 cccgtatctttaaccctccctgttaagttgtctctccagaatacaactgttaacta 60
Db 150524 CCCGTATCTTTAAACCCCTTGTTAACTTGTCTCTCCAGAAACCGTAAACCTA 150465

61 caaatgtcttccaatgagacccacagatgagtcacatgactaagatcacacgtgagcc 120
Db 150464 CAAATGTTCTTCAATGAGACACACAGATGAGTCCATGACTAAGATCCACCGTGGACC 150405

121 ctgagccgagctcgtagaccatgctccagatgtaatgacatggaagcaccctccag 180
Db 150404 CTGAGCCGAGCTCGTACGCCATGCTCCGATGTATGACATTGAAAGCACCCCTCCGAG 150345

181 gaaatcctaactgacacacccctactatgcccacatcagcggagacagttgagcgt 240
Db 150344 GAAATCTCAACTGACACACCCCTACTATGCCCCCAATTACCGGACAGTTAGAGCGGT 150285

241 catcagcacaacctcccaacagacactgggtttctcgttgaagaggggagctgagac 300
Db 150284 CATCAGCCAACTCCCAACAGACACTTGGGTTTCTCTGTGAGAGGGGAGCTGAGAC 150225

301 aggaactagctggaattcctatagccaagaagaatccctaaagcctagctggagagtgact 360
Db 150224 AGGACTAGCTGGATTTCTTAGGCCAACGAAATCCTTAAGCCTAGCTGGAGAGTGACT 150165

361 gcatccaccttaaacatgagtggtctgcaacttagctcaacccgacacatcagaagctc 420
Db 150164 GCATCCACCTTAAACATGGGGCTTGCACTTACTACACCCGACCAATCAGAGACTC 150105

421 actaaatgctaataatagcaaaaatagaggttaagaataagcaatcatctatgctgt 480
Db 150104 ACTAAATGCTAATTAGCAAAAATAGAGGTAAAGAAATAGCAATCATCTATTGCTGT 150045

481 agagcacagcggaggaagcaagaatcgagatataaaccaggaactcgagcggcgaag 540
Db 150044 AGAGCACAGCGGAGGAGCAAGATCGGATATTAACCCAGGCACTGAGCGGCAACGG 149985

541 caaccccttggtgtccctcccttctgtaggggctgtttcatctatctatctc 600
Db 149984 CAACCCCTTGTGTCCCTCCCTTGTATGGGGCTGTGTTTCACCTTATTTCACTCT 149925

601 attaatcttgcaactgaaaaaataaaaaa 635
Db 149924 ATTAATCTTGCAACTGAAAAAATAAAAAAGAAAA 149890

RESULT 4
AC022171 210336 bp DNA 1linear HTG 26-MAR-2001
LOCUS Homo sapiens chromosome 18 clone RP11-407618, WORKING DRAFT
DEFINITION AC022171.2 unordered pieces.
ACCESSION AC022171 GI:13270574
VERSION AC022171.18 GI:13270574
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F., Boguslavsky, L., Boukhaltier, B., Brown, A., Burkett, G., Castle, A., Chepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Deaellano, K., Dewar, K., Domino, M., Doyle, M., Fenesor, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lehocsky, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Pierre, N., Pisanl, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tefaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 210336)

AUTHORS

Bruno, D., Conn, L., Dela Rosa, M., Federspiel, N., Foreman, P., Glukhova, S., Hansen, N., Hyman, R., Mao, J., Marathe, R., Morehouse, A. J., Oefner, P., Palm, C. J., Ramirez, D., Wilhelm, J., Yu, S., and Davis, R. W.

TITLE

Direct Submission
Submitted (26-JAN-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA

COMMENT

On Mar 10, 2001 this sequence version replaced gi:13122770.

----- Genome Center

Center: Stanford DNA Sequencing and Technology Development

Center

Center code: SDSTDC

Web site: <http://sequence-www.stanford.edu/group/human/>

Contact: hum-info@sequence.stanford.edu

----- Project Information

Center project name: 758

Center clone name: RP11-407C18

----- Summary Statistics

Sequencing Vector: M13mp18; X02513; 97% of reads

Sequencing Vector: plasmid; plasmid_accession: 3% of reads

Chemistry: Dye-primer; 10% of reads

Chemistry: Dye-terminator Big Dye; 88% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 208312 bases at least Q40

Consensus quality: 208877 bases at least Q30

Consensus quality: 209157 bases at least Q20

Insert size: 199762; agarose-fp

Insert size: 210236; sum-of-contigs

Quality coverage: 8.7x in Q20 bases; agarose-fp

Quality coverage: 8.3x in Q20 bases; sum-of-contigs.

* NOTE: This is a working draft sequence. It currently

* consists of 2 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 66468: contig of 66468 bp in length

* 66469 66568: gap of unknown length

* 66569 210336: contig of 143768 bp in length.

Location/Qualifiers

1..210336

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="18"

/clone="RP11-407C18"

/clone.lib="RRCI human BAC library 11"

1..66468

/note="assembly_name:Contig44

clone_end:17"

66569..210336

/note="assembly_name:Contig45

clone_end:SP6"

BASE COUNT 67675 a 37228 c 37515 g 67816 t 102 others

ORIGIN

Query Match 99.5%; Score 631.8; DB 2; Length 210336;

Best Local Similarity 99.7%; Pred. No. 1.2e-190;

Matches 633; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ccctgtaccttaacctctgttaagttgtctctccagaatcaaaactgtaaaacta 60

Db 58176 CCCTGTACTTAACTCCTCTGTTAGCTTCTCTCCAGATCAAAACGTAACATA 58235

QY 61 caaatgtcttcaaatgagacacagatgagtcacatgactaagaatccaccgtgacc 120

Db 58236 CAAATGCTTCTTCAAAATGAGACACAGATGAGTCATGATCAATGATCACCCTGACCC 58295

QY 121 ctgagccggcctgtgacgcacatgctcgcagatgttaatgacatgttaaggaccctccag 180

Db 58236 CTGACCGCGCTGCTAGCCCATGCTCCCATGTTAATGACATTTGAAGCACCCTCCGAG 58335

QY 181 gaaatcctaactgcacaacacctactactatgcaccaatcaagcggagagttagaagcgt 240

Db 58356 GAAATCTCAACTGACACAAACCCCTACTATAGCCCATATTCAGGGGGAAGCAGTTAGACCGGT 58415

QY 241 catcagccaacctcccacaagcaacttgggtttcttcctgtgagaggggagactagaagac 300

Db 58416 CATCAGCCAACTCCCTCCCAACAGCAGCTTGGGTTTCTCTTGGAGGGGGGACTGAGAAC 58475

QY 301 aggaactagctgaattctcttaagcaagaagaatccccaagccttagcctggagaaagtac 360

Db 58476 AGGACTACTGTGATTTCTTCTAGGCCAAGCAGATCTCTTAAGCTTACTGCTGGAGAGTGACT 58535

QY 351 gcatccacctctaaacatgaggctgtgcaacttagctcacaccgcgaacatagaagctc 420

Db 58536 GCATCCACCTCTTAACATGAGGGCTTGCAACTTACTGACACCCGACCATCAGACAGCTC 58595

QY 421 actaaatgctaataataggcaaaaataggaggtlaagaaatagccaatcatctatgtcgt 480

Db 58596 ACTAAATGCTAATATGACCAAAAATAGAGGTAAAGAAATAGCCAAATCATCTATTGCGCTG 58655

QY 481 agagacagcgggagggagcaagagatcgagatataaacccagcattcgagccggaacgg 540

Db 58656 AGAGCAGACGGGAGGAGGACAAAGATCGGATTAATTAACCAAGCATTGACGCGGACACGG 58715

QY 541 caaccctcttgggtccctccctctgtatgagcgctctgtttcactctatctactct 600

Db 58716 CAACCCCTTTGGGTCCTCCCTCTTGTATGGGCGCTCTGTTTCACTCTATTCTACTCT 58775

QY 601 attaaatcttgcactgaaaaaataaaaaa 635

Db 58776 ATTAATCTTGCACACTGAAAAAAGAAAAA 58810

RESULT 5 AC010778 148724 bp DNA linear HTG 30-MAR-2000

LOCUS AC010778/c Homo sapiens clone RP11-2N15, WORKING DRAFT SEQUENCE, 8 unordered

DEFINITION pieces.

AC010778

VERSION AC010778.3 GI:7341824

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE human.

ORGANISM

Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 148724)

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE Homo sapiens, clone RP11-2N15

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 148724)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,

Baldwin, J., Barne, N., Beckerly, R., Boguski, L., Boukhalter, B.,

Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,

Cooke, P., Deavellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,

Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,

Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,

Lehoczky, J., Lieu, C., Locke, R., Macdonald, P., Marquis, A.,

McMan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,

Morrow, K., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,

Peterson, J., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Teste, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,

Wymann, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (22-SEP-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 30, 2000 this sequence version replaced gi:6087973.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Em., EMBL, SW., SWISSPROT, Tr., TREMBL, Wp., WORMEP, Information on the WORMEP database can be found at <http://www.sanger.ac.uk/Projects/C.legans/wormep> This sequence was generated from part of bacterial clone contigs of human chromosome X constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/ChrX>

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP6-198C4 is from the library RPCI-6 constructed by the group of Pieter de Jong. For further details see <http://www.choil.org/bacpac/home.htm>

VECTOR: pPAC4

IMPORTANT: This sequence is not the entire insert of clone RP6-198C4. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

FEATURES

source 1..8339
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q13.3-21.2"
/clone="RP6-198C4"
/clone.lib="RPCI-6"

repeat_region 1..5641 repeat: matches 368..5930 of consensus"
repeat_region 5635..5673 repeat: matches 7419..7457 of consensus"
repeat_region 5664..5955 repeat: matches 8244..8523 of consensus"
repeat_region 5956..6136 repeat: matches 1..239 of consensus"
repeat_region 6135..6280 repeat: matches 239..324 of consensus"

BASE COUNT 2649 a 1846 c 1726 g 2118 t
ORIGIN

Query Match 91.2%; Score 579.2; DB 9; Length 8339;
Best Local Similarity 94.8%; Pred. No. 4.4e-174;
Matches 599; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```

1 cccctactcttaactctctgttaagtgtctcttcagagataaactgtaacta 60
5664 cccctactcttaactctctgttaagtgtctcttcagagataaactgtaacta 5723
OY 61 caaatgtcttcaaatgagacacagatgagatcactgaacacgctgaccc 120
Db 5724 CAAATGTTCTTCAATGAGACCCCAAGATGCATGACATAAACCACCGCGATCC 5783
OY 121 ctgagcggcctgtcagccatgctcgaatgtaagcaatgaaagccctccgag 180
Db 5784 CTGGACCGACCGCTGAGCCCATGCTCCAGTGTATGATGCAAGGACACCCCTCCGAG 5843
OY 181 gaaatctaaactgagacacccctactatgcccccaattcagcgggagagatgaagcggt 240
Db 5844 GAAATCTAACTGACAAACCCCTACTATGCCCCCAATTCAGCAGGAGCACTTAAAGCAGT 5903
OY 241 catcagcaaacctcccaacagacacttggtttctctgttgagaggggagctggaagac 300
Db 5904 CATGAGCAACTTCCCAACAGCACTTGGGTTTCTCTTGAAGAGGGGTACTGAGAGAC 5963
OY 301 aggaactagctgattctcctagggcaagaaatccctaaagcctagctgggaaggtgact 360
Db 5964 AGGACTACTGATTTCTAGGCAACTAAGAAATCCCTAAGCCTAGCTGGGAAGGTGACC 6023
OY 361 gcatccactctcaaacatgagggtctgtcgaactagctcacccgccaatcagaagatc 420

```

```

Db 6024 ACACCCACCTTTAAACATGCGGCTTGCAACTTACCTACACCCCAACCAATCAGAGAGCTC 6083
OY 421 actaaatgctaattaggcaaaaatagagagtaaaagaatagccaatcattcctg 480
Db 6084 ACTAAATGCTAATTATGACCAAAAACAGAGATGAAGAAATAGCCAAATCATCTATTGCTG 6143
OY 481 agagacacgcggaagagcaaaagatcgggataaaacccaagcattcgagccggaag 540
Db 6144 AGAGCAGAGCTGAGAGGACAAAGATCGGAGATTAACCAAGCATTCTGGCCAGACAG 6203
OY 541 caaccctcttggtgcccctccctctgtatagggcgctgttttactatattact 600
Db 6204 CAACCCCTTGGTGGTCCCTCCCTTGTATGGAGACTGTGTTGACTATTTACCTCT 6263
OY 601 attaaacttgcaactgaaaaaataaaaaa 632
Db 6264 ATTAATCTTGCAACTGAAAAAATAATGCAN 6295

```

RESULT 7
AC073626/c 136901 bp DNA linear PRI 09-MAY-2001
LOCUS AC073626
DEFINITION Homo sapiens BAC clone RP11-95P9 from 7, complete sequence.
ACCESSION AC073626
VERSION AC073626.7 GI:12863221
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 136901)
Sulston,J.E. and Waterston,R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
2 (bases 1 to 136901)
Nguyen,C., Haakenson,B. and Bauer,H.
The sequence of Homo sapiens BAC clone RP11-95P9
Unpublished
3 (bases 1 to 136901)
Waterston,R.H.
Direct Submission
Submitted (27-JUN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 136901)
Waterston,R.H.
Direct Submission
Submitted (16-FEB-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 136901)
Waterston,R.
Direct Submission
Submitted (09-MAY-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 16, 2001 this sequence version replaced g1:11597125.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapliens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0095P09

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate

chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRF/CHR7>, send <mailto:sgreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenio, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://daccpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is CTD-2023N18, 200 bp overlap: the clone sequenced to the right is CTA-250D13, 200 bp overlap. Actual start of this clone is at base position 127066 of CTD-2023N18; actual end is at base position 26937 of CTA-250D13.

There are polymorphic base differences between the clone RP11-95P9 and stolen data from a redundant clone RP11-636N5. The region from base position 29435 to 29729 consists of an insertion fragment from RP11-636N5, which was originally found deleted RP11-95P9.

FEATURES

source

1. 136901
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-95P9"
/clone_1lb="RPCT-11"
545..1157
/rpt_family="L1"
1195..1360
/rpt_family="L1"
1530..1791
/rpt_family="L1"
2763..3320
/note="similar to EST AA564655 (NID:g2336294) nj25902.sl"
2972..3112
/rpt_family="MERL_type"
3118..3308
/rpt_family="MERL_type"
6682..6885
/rpt_family="MIR"
7038..7388
/rpt_family="MaLR"
7433..7802
/rpt_family="MaLR"
7811..7938
/rpt_family="(TA)n"
7945..7994
/rpt_family="AT_rich"
8005..8064
/rpt_family="(TATATG)n"
8067..8206
/rpt_family="MaLR"
9047..9246
/rpt_family="MaLR"
9227..9306
misc_feature

/note="similar to EST AU035546 (NID:g3718549)"
9289..9399
/rpt_family="MaLR"
9412..9494
/rpt_family="MaLR"
9495..9854
/rpt_family="MaLR"
9855..9928
/rpt_family="MaLR"
10530..10814
/rpt_family="Alu"
10815..10909
/rpt_family="(CATG)n"
12990..13351
/note="similar to EST AV755789 (NID:g10913637)"
12990..13414
/note="similar to EST AW955488 (NID:g8145171)"
13025..13434
/note="similar to EST AV756558 (NID:g10914406)"
13026..13433
/note="similar to EST AV710618 (NID:g10729924)"
13026..13414
/note="similar to EST BE888689 (NID:g10345245)"
13026..13425
/note="similar to EST AW975063 (NID:g8166267)"
13026..13433
/note="similar to EST AV717668 (NID:g10814820)"
13026..13433
/note="similar to EST AV762499 (NID:g10920347)"
13026..13433
/note="similar to EST AV701025 (NID:g10717355)"
13026..13222
/note="similar to EST BF348560 (NID:g11296155)"
13026..13411
/note="similar to EST AV755769 (NID:g10913617)"
13026..13433
/note="similar to EST AV760867 (NID:g10918715)"
13026..13405
/note="similar to EST BE277509 (NID:g9152479)"
13026..13408
/note="similar to EST AA532593 (NID:g2276847) nj54e08.sl"
13026..13357
/note="similar to EST BE907116 (NID:g10400459)"
13028..13430
/note="similar to EST BF031653 (NID:g10739345)"
13031..13434
/note="similar to EST A1879643 (NID:g5553692) au56e10.x1"
13031..13433
/note="similar to EST AW157724 (NID:g6229125) au86a05.x1"
13032..13224
/note="similar to EST AA608523 (NID:g2456951) ae53b04.sl"
13035..13171
/note="similar to EST A1476367 (NID:g4329412) cm21d12.x1"
13038..13411
/note="similar to EST BF237771 (NID:g11151690)"
13040..13404
/note="similar to EST BE393937 (NID:g9334762)"
13053..13434
/note="similar to EST A1879261 (NID:g5553310) au56e10.y1"
13119..13435
/note="similar to EST BE710307 (NID:g10098572)"
13182..13413
/note="similar to EST BF034781 (NID:g10742493)"
14398..14434
/rpt_family="(TAA)n"
14453..14659
/rpt_family="MIR"
15131..15248
/rpt_family="MER94"
15255..15456
/rpt_family="MaLR"
15457..15494
/rpt_family="AT_rich"

repeat_region 15515..15552
/rpe_family="AT_r1ch"

Query Match 89.9%; Score 571; DB 9; Length 136901;
Best Local Similarity 93.7%; Pred. No. 3.4e-171;
Matches 595; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

```

OY 1 cccgtatcttaacctccctgttaagttgtctctccagaatcaaaactgtaaaacta 60
    |||||
DB 95677 cccgtatcttaacctccctgttaagttgtctctccagaatcaaaactgtaaaacta 95618
OY 61 caaatgtcttccaatgagagcagatgagtcacatgacaaatccacgctgagacc 120
    |||||
DB 95617 caaatggttttccaatgagagcagatgagtcacatgacaaatccacgctgagacc 95558
OY 121 ctggagcggcgtctgtagcccatgctcgaatgttaatgaatgagagccctcccgag 180
    |||||
DB 95557 ctggaatggcctgtagcccatgctcgaatgttaatgaatgagagccctcccgag 95498
OY 181 gaaatctcaactgcaaacccctactatgcccacaaatccagcggagagcagtcgct 240
    |||||
DB 95497 gaaatctcaactgcaaacccctactatgcccacaaatccagcggagagcagtcgct 95438
OY 241 catcagccaacctcccccaacagcaacttggtttccctgttgagaggggagctgagagac 300
    |||||
DB 95437 catcagccaacctcccccaacagcaacttggtttccctgttgagaggggagctgagagac 95378
OY 301 aggaactagctgagttcttcctcgaagcgaagatccctcgaagctagctgagaggtgact 360
    |||||
DB 95377 aggaactagctgagttcttcctcgaagcgaagatccctcgaagctagctgagaggtgact 95318
OY 361 gcatcacaacctcaaacatgagggctgcaactagctcacaccgccaatcagagagctc 420
    |||||
DB 95317 gcatcacaacctcaaacatgagggctgcaactagctcacaccgccaatcagagagctc 95258
OY 421 actaaatgctcaatagagcaaaaatagaggtgaagaagaatagcaatcactatgctgctg 480
    |||||
DB 95257 actaaatgctcaatagagcaaaaatagaggtgaagaagaatagcaatcactatgctgctg 95198
OY 481 agagacagcggagaggaagcaagatcgggatalaaaccagagcattcgagccgcaacg 540
    |||||
DB 95197 agagacagcggagaggaagcaagatcgggatalaaaccagagcattcgagccgcaacg 95138
OY 541 caaaccccttggtgctccctccttgatgagggcgctgctgttcaactatctactct 600
    |||||
DB 95137 caaaccccttggtgctccctccttgatgagggcgctgctgttcaactatctactct 95078
OY 601 attaaatctgcaactgaaaaaataaaaaa 635
    |||||
DB 95077 attaaatctgcaactgaaaaaataaaaaa 95043

```

RESULT 8
AL583805 89728 bp DNA linear PRI 29-SEP-2001
LOCUS Human DNA sequence from clone RP11-134K1 on chromosome 9, complete
DEFINITION
Sequence.
ACCESSION AL583805 AC051658
VERSION AL583805.7 GI:15865009
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 89728)
AUTHORS Clark, G.
TITLE Direct Submission
JOURNAL Submitted (29-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Oct 2, 2001 this sequence version replaced gi:14702155.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; information on the WormPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr9>
RP11-134K1 is from the library RPEC1-11.1 constructed by the group of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-134K1 it may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true right end of clone RP11-134K1 is at 89728 in this sequence. The true right end of clone RP11-60C15 is at 2000 in this sequence.

FEATURES

source
1..89728
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-134K1"
/clone_1lb="RPEC1-11.1"
BASE COUNT 29463 a 17925 c 16432 g 25908 t
ORIGIN

Query Match 88.9%; Score 564.6; DB 9; Length 89728;
Best Local Similarity 94.2%; Pred. No. 3.5e-169;
Matches 598; Conservative 0; Mismatches 34; Indels 3; Gaps 1;

```

OY 1 cccgtatcttaacctccctgttaagttgtctctccagaatcaaaactgtaaaacta 60
    |||||
DB 83434 cccgtatcttaacctccctgttaagttgtctctccagaatcaaaactgtaaaacta 83490
OY 61 caaatgtcttccaatgagagcagatgagtcacatgacaaatccacgctgagacc 120
    |||||
DB 83491 caaatggttttccaatgagagcagatgagtcacatgacaaatccacgctgagacc 83550
OY 121 ctggagcggcgtctgtagcccatgctcgaatgttaatgaatgagagccctcccgag 180
    |||||
DB 83551 ctggagcggcgtctgtagcccatgctcgaatgttaatgaatgagagccctcccgag 83610
OY 181 gaaatctcaactgcaaacccctactatgcccacaaatccagcggagagcagtcgct 240
    |||||
DB 83611 gaaatctcaactgcaaacccctactatgcccacaaatccagcggagagcagtcgct 83670
OY 241 catcagccaacctcccccaacagcaacttggtttccctgttgagaggggagctgagagac 300
    |||||
DB 83671 catcagccaacctcccccaacagcaacttggtttccctgttgagaggggagctgagagac 83730
OY 301 aggaactagctgagttcttcctcgaagcgaagatccctcgaagctagctgagaggtgact 360
    |||||
DB 83731 aggaactagctgagttcttcctcgaagcgaagatccctcgaagctagctgagaggtgact 83790
OY 361 gcatcacaacctcaaacatgagggctgcaactagctcacaccgccaatcagagagctc 420
    |||||
DB 83791 gcatcacaacctcaaacatgagggctgcaactagctcacaccgccaatcagagagctc 83850

```

```

Oy 421 actaaagtgtaattagcaaaaataaggagtaagaataagccaatcattatgctg 480
Db 83851 ACTAAATGCTAATTAGCAAAAAACGAGAGTAAGAAATAGCAATCATCTATTGCTG 83910
Oy 481 agagcaagagggagggaggaagatcggatataaccggcatttcgagccggcaacgg 540
Db 83911 AGAGCAGACAGAGAGGAGCAAGATGAGATATTAACCCAGGCAATTCGAGCGGCAACAG 83970
Oy 541 caaccccttggtggtccccccttgatggcgctcgttttctactatctactct 600
Db 83971 CAACCCCTTGGTGCTCCCTCCCTTTATGGAGCTCTGTTTTCATCTATTTCACCTCT 84030
Oy 601 attaaatctgcactgtaaaaaaataaaaaa 635
Db 84031 ATTAATCTGTCGCAACTCAAAAAAATAAAAAA 84065

RESULT 9
AL451130/c
ACCUSSION
DEFINITION
AL451130 212096 bp DNA linear HTG 08-FEB-2002
PROGRESS *** 9 unordered pieces.
AL451130 AC011776
AL451130.7 GI:18643763
HTG: HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (sites)
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (07-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk; Clone requests: clonerequests@sanger.ac.uk
On Feb 10, 2002 this sequence version replaced gi:18151469.
Draft Sequence Produced by Whitehead Institute/MIT Center for
Genome Research, 320 Charles street,
Cambridge, MA 02141, USA
http://www-seq.wi.mit.edu
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BA10621
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M77815; 5% of reads
Sequencing vector: plasmid; L08752; 40% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 206856 bases at least Q40
Consensus quality: 208786 bases at least Q30
Consensus quality: 210158 bases at least Q20
Insert size: 211296; sum-of-ctnigs
Insert size: 183186; 9.1% error; agarose-fp
Quality coverage: 9.78x in Q20 bases; sum-of-ctnigs Quality
Coverage: 12.66x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2459: contig of 2459 bp in length
* 2460 2559: gap of 100 bp
* 2560 4643: contig of 2084 bp in length
* 4644 4743: gap of 100 bp
* 4744 6823: contig of 2080 bp in length

```

```

* 6824 6923: gap of 100 bp
* 6924 9152: contig of 2229 bp in length
* 9153 9252: gap of 100 bp
* 9253 198727: contig of 189475 bp in length
* 198728 198827: gap of 100 bp
* 198828 202432: contig of 3605 bp in length
* 202433 202532: gap of 100 bp
* 202533 206117: contig of 3585 bp in length
* 206118 206217: gap of 100 bp
* 206218 209532: contig of 3315 bp in length
* 209533 209632: gap of 100 bp
* 209633 212096: contig of 2464 bp in length.
FEATURES
source
1..212096
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-10621"
/clone_11b="RPC1-11.1"
1..2459
/note="assembly_fragment:00156"
2560..4643
/note="assembly_fragment:00601"
4744..6823
/note="assembly_fragment:01422"
6924..9152
/note="assembly_fragment:01529"
9253..198727
/note="assembly_fragment:02535"
198828..202432
/note="assembly_fragment:03097"
202533..206117
/note="assembly_fragment:03385"
206218..209532
/note="assembly_fragment:04246"
209633..212096
/note="assembly_fragment:04283"
BASE COUNT 63105 a 38564 c 41164 g 68462 t 801 others
ORIGIN
Query Match 88.9%; Score 564.6; DB 2; Length 212096;
Best Local Similarity 94.2%; Pred. No. 4.2e-169;
Matches 598; Conservative 0; Mismatches 34; Indels 3; Gaps 1;
Oy 1 cccgtatctttaaaccctctctgttaagtgtgtctcttccaaatcaaacgttaacta 60
Db 113775 CCGGTATCTTTAACTCTCTGTTAAGTGTCTCTTCCAGAAATCGAAGCTGTAA--TA 113719
Oy 61 caattgtcttcaaatgagacacagatgagatccatgactaaatccacgtgagacc 120
Db 113718 TATATGTTCTTCAATATGAGACACAGATGCAATGATTAAGATTAACCGGAGACC 113659
Oy 121 ctgagaccgacctgtagccatgctcgcagtgtaatgacattgaaagcaccctccgag 180
Db 113658 CTGAGACCGGCTGCTAGGCCATGCTCGATGTAAACGACATTGAAGGACACCCCTCTAG 113599
Oy 181 gaaatctaaactgcaaacccctactatgcccccaattcagcgagggaaggttagagcggt 240
Db 113598 GAAATCTCACTGACCAACCAACTACTATGCCCAATTCAGCAGGAAGCAGTAGAGCGGT 113539
Oy 241 catcagcaactccccaagacagacttggtttctctgttagagagggagctgagagac 300
Db 113538 TGTACGCAACCTCCCAACGACACTTGGGTTTCTTTAGAGACGGGAGCTAGAGAC 113479
Oy 301 agagctagctgattctctagcaacgaagaatcccttaagccttagctggaaggtgact 360
Db 113478 AGGACTAGCTGGAATTTCTTAGCGGAATTAAGAAATCCCTTAAGCTGAGGAGGACT 113419
Oy 361 gcatccactttaaacaaggggcttgcaacttagctacaccgacccaataagagagtc 420
Db 113418 GCATCCACTTTAAACACAGGGGCTTGCACACTTAGCTCACACCACCAACATAGAGAGCTC 113359

```

OY	421	actaaagcgcattatgagcaaaaacaggaggtgaagaataaacatcatctatcgccg	480
Db	113358	ACTAAATGCTAATTAGGCACAAAACAGCAGTAAGAATAATAGCCATCATCTATTGGCCCG	113299
OY	481	agagcacagcgaggaggaacaagatcgggatataaaaccaggcaltcgagccgcaacg	540
Db	113298	AGAGCACAGCAGGAGGAGCAAGAGATGTAAACCCAGGCAATTTCGAGCCGGCAACAG	113239
OY	541	caacccccttgggtcccccttccttgatagtgggcctctgttttcaactctattcaacct	600
Db	113238	CAACCCCCTTGGGGTGCCCTCCCTTTGAATGGAGGACCTGTGTTTCATCTATTTCACCTCT	113179
OY	601	attaactctgcaactgaaaaaanaaaaaaaaaa	635
Db	113178	ATTAAATCTCTGCACTGCCAAAAAANAANAAAAA	113144
RESULT	10		
ACOT2023/c			
LOCUS			
DEFINITION	AC072023	175040 bp DNA linear HTG 31-JAN-2002	
ACCESSION	AC072023		
VERSION	AC072023.4		
KEYWORDS	HTG: HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 175040)		
AUTHORS	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alshrocks,S.L., Amaralunge,H.C., Are,J.R., Ayale,M., Banks,T., Barbatella,J., Benton,J., Bimarge,K., Blankenburg,K., Bonnin,D.P., Bouck,J.J., Bowls,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Butney,C., Burch,P., Burkett,C., Butrell,K.B., Byrd,N.C., Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., David,M.L., Davis,C., Davy-carroll,L., Dedertich,D.A., Delaney,K.R., Delgado,O., Deng,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Eamhart,C., Edgar,D., Edwards,C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franze,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorell,J.D.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsl,F., Howard,S., Huber,J., Huliyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsom,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kralovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichteberg,O., Lieu,C., Liu,J., Liu,W., Louisedge,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhinney,E., Mcloed,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Mosser,M., Neel,D.J., Newton,S., Ng,N., Nguyen,A., Nguyen,N., Nuygen,N., Nickerson,E., Nwokkenko,S., Ogun,M., Okunodu,G., Oregunye,N., Oyiedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Petere,L., Pickens,R., Picea,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtali,N., Sisson,I., Sodergren,E., Sonakke,T., Sparks,A., Stanley,H., Stone,H., Sulton,A., Svatek,A., Tabot,P., Tameris,K., Tamerisa,K., Tang,H., Taney,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Umani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wellington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.		
TITLE	Unpublished		
JOURNAL	Direct Submission		
REFERENCE	2 (bases 1 to 175040)		

```

AUTHORS      Morley,K.C.
TITLE        Direct Submission
JOURNAL      Submitted (07-JUN-2000) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
COMMENT      On Dec 10, 2001 this sequence version replaced gi:9438656.
              -----
              Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
              Project Information
              Center project name: HBEX
              Center clone name: RP11-3j2
              -----
              Summary Statistics
findhaplist
Consensus quality: 177002 bases at least Q40
Consensus quality: 181942 bases at least Q30
Consensus quality: 185128 bases at least Q20
Estimated insert size: 180460; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-ip estimation
Quality coverage: 5.1x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This is a "working draft" sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
  be preserved.
    1       47110: contig of 47110 bp in length
    *         47111       47210: gap of unknown length
    *         47211       80061: contig of 32851 bp in length
    *         80062       80161: gap of unknown length
    *         80162       112177: contig of 32016 bp in length
    *         112178     112277: gap of unknown length
    *         112278     140321: contig of 28044 bp in length
    *         140322     140421: gap of unknown length
    *         140422     156477: contig of 17226 bp in length
    *         15648      157747: gap of unknown length
    *         157748     168444: contig of 10697 bp in length
    *         168445     175040: gap of unknown length
    *         168545     175040: contig of 6496 bp in length.
FEATURES             Location/Qualifiers
     source            1..175040
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="3"
                        /clone="RP11-3j2"
BASE COUNT          53880 a 32737 c 32187 g 55620 t 616 others
ORIGIN
Query Match           88.2%; Score 560.2; DB 2; Length 175040;
Best Local Similarity 93.8%; Pred. No. le-167;
Matches 594; Conservative 0; Mismatches 38; Indels 1; Gaps 1
OY   1 cccgtatccttaaacctccttgtaaagtttgccttcctccagaataaaactgtaaaacta 60
      |||||
Db 123368 CCGTGATCTTTAAACCTCCTTGTTAAGTTTGTCCTCCAGAAATCAAAATGTTAAACTA 123309
OY   61 caaatgtcttcaaatgatgaacacacagatggatcatagacataaagtattccaccgtagccc 120
      |||||
Db 123308 CAATATGGGCTTTAAATGAAGGCCCAAGATGCAGTCATGATCAAGATCACACAGACC 123249
OY   121 ctgacacgacctgctgaccatctgctcgatcgttaagtacaatgaaggacccccctccgag 180
      |||||
Db 123248 CTGAGACTGGCGCTCTAAGCCCATGATCCGATGTTTAATGACATCAAAAGCACCCCTCCCAAG 123189

```

QY 181 gaattcacaactgcacacccctactatgcgcccaatcagcgggaagcagttgaagcgt 240
|||||
Db 123188 GAATCTCACTGCGCCCAACCCCTACTATGCCCAATTCAGAGAGAACAGTATAGAGAGT 123129
QY 241 catcagcacaactcccccacacagcacttgggtttctctgttgaagggggagactgaagac 300
|||||
Db 123128 CATTTGGCCCAACCTCCCAACAGCAGCTTGATCTCTGTTGAGAGGGGGGAGTGAAGAT 123069
QY 301 aggaactagctggatcttcttgcgcgaagaaatccctaagcctaagcttgggaagttact 360
|||||
Db 123068 AGGACTAGCTGGATTTCTTAGGCCAACTAATCCCTAAGCTTAGTGGAAGGTGACC 123009
QY 361 gcatccacctcaacataggggcttgcacactcaccacccagcaatcaagagactc 420
|||||
Db 123008 GCATCCACCTTTAAACACGGGGGCTCGCACTTACTCACCACCAATCAGAGAGCTC 122949
QY 421 actaaatgtcctaattagcaaaaaataaggaagtaagaataagccaatcatctatgtcctg 480
|||||
Db 122948 ACTMAATGCTAATTAGGCAAAAACAGAGGTAAA-TAATAGCCCAATCATCTATTGCTG 122890
QY 481 agagcacagcgggagggagacaagatcgggatataaccacgggacttgaagccggaagc 540
|||||
Db 122889 AGACACAGCGGGAGGAGCAAGAGGATATAAACCAAGGCAATTCAGAGCTGCAATGG 122830
QY 541 caaccccttgggtccctccctcttctgtatggcgctctgtttcactctatctactct 600
|||||
Db 122829 CAACCCCTTTGGGTGCCCTTCCCTTGTATGGGAGCTCTGTTTCTACTCTATTCTACTCT 122770
QY 601 attaaacttgcacactgaaaaaataaaaaa 633
|||||
Db 122769 ATTAATCTTGCAACTGCAAAAAAATAAAAAA 122737

RESULT 11
AP001538/c 174019 bp DNA linear PRI 25-MAR-2000
LOCUS Homo sapiens genomic DNA, chromosome 21q21.1-q21.2 clone:B853K11,
1L56-APP region, complete sequence.
ACCESSION AP001538
VERSION AP001538.1 GI:7328982
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:B853K11.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 174019)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 174,019 genomic DNA of 21q21.1-q21.2
Published Only in Database (2000) In press
2 (bases 1 to 174019)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (23-MAR-2000) to the DDBJ/EMBL/GenBank databases.
Masahira Hattori, The Institute of Physical and Chemical Research
(RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1
Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gs.c.riken.go.jp, UR: http://ngp.gsc.riken.go.jp/,
Tel:81-42-778-9923, Fax:81-42-778-9924)
Location/Qualifiers
1. 174019
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/clone="B853K11"
/map="21q21.1-q21.2"
BASE COUNT 53303 a 31416 c 32248 g 57052 t
ORIGIN

Query Match 87.7%; Score 556.6; DB 9; Length 174019;
Best Local Similarity 93.7%; Pred. No. 1,5e-166;

Matches 580; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 1 cccgtatctttaaaccctccctgttaagttgtctcttccagaatcaaaactgttaacta 60
|||||
Db 36710 CCCGTATCTTTAAACCTCTGTTAACTTGTCTTCCAGAAACAAACGTAAACTA 36651
QY 61 caatttcttcaatggagcaccagatgagtcatactgaataaccacgtggagcc 120
|||||
Db 36650 CAATTAAGTTCTTCAATATGGGGGCCCAATGACATGACCTAGACTAAGATCCACCGGGGCC 36591
QY 121 ctggaccgacctctgtagcccatgtctcgatgttaatgacaltgaaggcacccctccgag 180
|||||
Db 36590 CTAGACTAGCGCTGCTAGCCCATGCTCGGATATTAATGACATCGAAGCACCCTCCCAAG 36531
QY 181 gaattcacaactgcacacccctactatgcgcccaatcagcgggaagcagttgaagcgt 240
|||||
Db 36530 GAAATCAACACTGCGCCCAACCCCTAATGCCCCCAATTCAGAGGAGAACAGTAAAGGTGT 36471
QY 241 catcagcacaactcccccacacagcacttgggtttctctgttgaaggggagcgaagac 300
|||||
Db 36470 CATTTGGCCCAACCTCCCAACAGCAGCTTGGGTTTCTGTTGAGGGGGGTRCTAGAGAC 36411
QY 301 aggaactagctggatcttcttgcgcgaagaaatccctaagcctaagcttgggaagttact 360
|||||
Db 36410 AGGACTAGCTGGATTTCTTAGGCCAACTAATCCCTAAGCTTAGTGGAAGGTGACC 36351
QY 361 gcatccacctcaacataggggcttgcacactcaccacccagcaatcaagagactc 420
|||||
Db 36350 GCATCCACCTTTTAAACATGGGGGCTCGCAACTTACTCTACACCAACAGTCAGAGACTC 36291
QY 421 actaaatgtcctaattagcaaaaaataaggaagtaagaataagccaatcatctatgtcctg 480
|||||
Db 36290 ACTMAATGCTAATTAGGCAAAAACAGAGGTAAAGAAATGCCAATCATCTATTGCTGCT 36231
QY 481 agagcacagcgggagggagacaagatcgggatataaaccagcatctgagccggaagc 540
|||||
Db 36230 AGACACAGCGGGAGGAGCAAGAGATGATGAATCCCAAGCTTCAAGCTGCAACAGG 36171
QY 541 caaccccttgggtccctccctcttctgtatggcgctctgtttcactctatctactct 600
|||||
Db 36170 CAACCCCTTTGGGTGCCCTTCCCTTGTATGGGAGCTCTGTTTCTACTCTATTCTACTCT 36111
QY 601 attaaacttgcacactgaa 619
|||||
Db 36110 ACTAATCTTGCAACTGCA 36092

RESULT 12
AP001674/c 340000 bp DNA linear PRI 30-MAY-2000
LOCUS Homo sapiens genomic DNA, chromosome 21q, section 18/105.
ACCESSION AP001674 AL163219 BA000005
VERSION AP001674.1 GI:7768666
KEYWORDS
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Seede,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M. and Schudy,A.
The DNA sequence of human chromosome 21. The chromosome 21 mapping
and sequencing consortium
Nature 405 (6784), 311-319 (2000)
JOURNAL MEDLINE
20289799
TITLE 2 (bases 1 to 340000)
Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Seede,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,

TITLE
JOURNAL

Patterson, D., Reichwald, K., Rump, A., Schillhabel, M., Schudy, A., Zimmermann, W., Rosenthal, A., Kudo, J., Shibuya, K., Kawasaki, K., Asakawa, S., Shintani, A., Sasaki, T., Nagamine, K., Mitsuyma, S., Antonarakis, S. E., Minoshima, S., Shimizu, N., Nordstiek, G., Hornischer, K., Barandt, P., Scharfe, M., Schoen, O., Desario, A., Reichelt, J., Kauer, G., Blocker, H., Ramser, J., Beck, A., Klages, S., Hennig, S., Krieselmann, L., Dagand, E., Wehrmeyer, S., Borzym, K., Gardiner, K., Nizetic, D., Francis, F., Lehnardt, H., Reinhardt, R. and Yaspo, M. L.

Direct Submission

Submitted (10-Apr-2000) to the DDBJ/EMBL/GenBank databases. The Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic Sciences Center, Human Genome Research Group * Institute of Molecular Biotechnology, Genome Analysis * Keio University School of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular Genetics (addresses see below)

On May 30, 2000 this sequence version replaced gi:7717271.

The chromosome 21 mapping and sequencing consortium consisting of

* RIKEN Genomic Sciences Center, Human Genome Research Group, * Sagemihara 228-8555, Japan,

* e.mail: hatorigsc.riken.go.jp/

* URL: http://hgp.gsc.riken.go.jp/

and

* Institute of Molecular Biotechnology, Genome Analysis, * Beutenbergstrasse 11, D-07745 Jena, Germany,

* e.mail: gscj-submit@genome.imb-jena.de

* URL: http://genome.imb-jena.de/

and

* Keio University School of Medicine, Molecular Biology, * Tokyo 160-8582, Japan,

* e.mail: nshimizu@mb.med.keio.ac.jp

* URL: http://www.dmb.med.keio.ac.jp/

and

* GBF, Dept. of Genome Analysis, * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail: info.genome@gbf.de

* URL: http://genome.gbf.de/

and

* Max-Planck Institute for Molecular Genetics, * Ihnestrasse 73, D-14195 Berlin, Germany,

* e.mail: info-chr21@molgen.mpg.de

* URL: http://chr21.rz-berlin.mpg.de/

AL163219: Submitted (10-Apr-2000).

Location/Qualifiers

1. 340000

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="21"

<1. 6149

/note="Accession No. AP001537"

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="21"

/map="21q21.1"

/clone="R749F18, 5' partial"

/clone_id="RPCT-11 BAC library"

<1. 165231

/note="Accession No. AP001538"

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="21"

/map="21q21.1"

/clone="R853K11, 5' partial"

/clone_id="RPCT-11 BAC library"

15. 444

/note="L1MA4A"

/rpt_family="LINE/L1"

/rpt_type="DISPERSED

555. 3531

/note="L1MA4A"

/rpt_family="LINE/L1"

/rpt_type="DISPERSED

repeat_region

3770. 3876

/note="L2"

/rpt_family="LINE/L2"

/rpt_type="DISPERSED

complement(5645. 6178)

/note="L1MA2"

/rpt_family="LINE/L1"

/rpt_type="DISPERSED

6181. 6534

/note="THE1B"

/rpt_family="LTR/MaLR"

/rpt_type="DISPERSED

8021. 8370

/note="THE1B"

/rpt_family="LTR/MaLR"

/rpt_type="DISPERSED

complement(8373. 10032)

/note="L1MA2"

/rpt_family="LINE/L1"

/rpt_type="DISPERSED

10213. 10520

/note="Alu/b"

/rpt_family="SINE/Alu"

/rpt_type="DISPERSED

15810. 15851

/note="(CA)n"

/rpt_family="Simple_repeat"

/rpt_type="TANDEM

complement(16628. 16697)

/note="MER53"

/rpt_family="DNA"

/rpt_type="DISPERSED

16638. 16715

/note="MER53"

/rpt_family="DNA"

/rpt_type="DISPERSED

18847. 19585

/note="L2"

/rpt_family="LINE/L2"

/rpt_type="DISPERSED

19795. 19917

/note="(TA)n"

/rpt_family="Simple_repeat"

/rpt_type="TANDEM

20023. 20166

/note="(TA)n"

/rpt_family="Simple_repeat"

/rpt_type="TANDEM

20289. 20448

/note="(TA)n"

/rpt_family="Simple_repeat"

/rpt_type="TANDEM

20467. 20542

/note="(TA)n"

/rpt_family="Simple_repeat"

/rpt_type="TANDEM

complement(21360. 22028)

/note="L1ME3"

/rpt_family="LINE/L1"

/rpt_type="DISPERSED

complement(22142. 22448)

/note="Alu/x"

/rpt_family="SINE/Alu"

/rpt_type="DISPERSED

complement(22519. 22891)

/note="MER52A"

/rpt_family="LTR/MER4-group"

/rpt_type="DISPERSED

22976. 23239

/note="L1MD3"

/rpt_family="LINE/L1"

/rpt_type="DISPERSED

complement(23244. 23519)

COMMENT

FEATURES
source

source

source

repeat_region

repeat_region

repeat_region

```

/ote="MER4C"
/rpl_family="LTR/MER4-group"
/rpl_type=DISPERSED
complement(23644..23661)
/ote="MER4C"
/rpl_family="LTR/MER4-group"
/rpl_type=DISPERSED
23925..25091
/ote="MER52A"
/rpl_family="LTR/MER4-group"
/rpl_type=DISPERSED
complement(25227..25672)
/ote="MER4A2"
/rpl_family="LTR/MER4-group"
/rpl_type=DISPERSED
complement(25682..25987)
/ote="AluY"
/rpl_family="SINE/Alu"
/rpl_type=DISPERSED
26744..26849
/ote="MER46B"
/rpl_family="DNA/MER2_type"
/rpl_type=DISPERSED
complement(26852..27629)
/ote="LTR17"
/rpl_family="LTR/Retroviral"
/rpl_type=DISPERSED
complement(27630..27921)
/ote="HERV17"
/rpl_family="LTR/Retroviral"

Query Match      87.7%; Score 556.6; DB 9; Length 340000;
Best Local Similarity 93.7%; Pred. No. 1,7e-166;
Matches 580; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 1 cccgtatcttaacctctctgttaagttgtctcttcgaagaatcaaatgttaaaacta 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27921 CCTGTATCTTTAACTCCCTGTTAACTTGTCTCTCCAGAAATCAAAACTGTAAACTA 27862

QY 61 caattgtcttcaaatgagacccagatgagtgatccatgaactaagaccgctgagacc 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27861 CAAATAGTTTTCATTAATGGGGCCCCCAGATGCATGCATGACTAAGATCCACGGGGGCC 27802

QY 121 ctgagacgctctgtagcccatgtctccgatgttaattgacattgaaggaccctccag 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27801 CTAGAGTGGCTCTGATGCCATGCTCGATATTAATGACATCGAAGCACCCTCCCAAG 27742

QY 181 gaaatctcaactgcaaacccctactatagcccaatcagcggggaagcagtagagcg 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27741 GAAATCACAACCTCCCAACCCCTACTATGCCCAATTCACGAGAACAGTATGAGCTGT 27682

QY 241 catcagccaacccctcccaagaacacttggtttctctgttgagaggggagactgaag 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27681 CATTTGGCACTTCCTCCCAAGACACTTGGTTTCTGTTGAGGGCGGTTACTGAGAGAC 27622

QY 301 aggaactagctgatttctctagccaagaagaatccctaagcctagctgaggaaagtact 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27621 AGAATCAGTGGATTCCTTAGGCTGACTAGATTCCTTAAGCTAGCTGGGAAGGTAC 27562

QY 361 gcatcaaccttaaacatggggttgtaacttagctcacaccgaccatcagaagatc 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27561 GCATTCACCTTTTAAACATGGGGCTCGCACTTAGCTCACACCACAGTCAGAGAGCTC 27502

QY 421 actaaatgtcttaattgagcaaaaataagagtaagaataatgacatcatctattgctgt 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27501 ACTTAAATGCTTAATTAAGCAAAACAGAGAGTAAAGAAATAGCCAAATCATCTATTGCTG 27442

QY 481 agagcacaagcgagagacaaagatcgagatatataaacccagcattcgaagcgcaacg 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27441 AGAGCACAAGCAGGAGGACAAAGATCGAGATGATAAACCCAGGCACTTCAAGCTGGCAACGG 27382

QY 541 caacccttgggtccctccctctgtatgagcgctgttcttactattactct 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 27381 CAACCCCTTTGGGTCCTCCCTGCTTGTATGAGAGCTGTGTTTCACTATTTCACTCT 27322
QY 601 attaatctgcaactgaa 619
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 27321 ACTTAAATCTTGCACTGCA 27303

RESULT 13
AP001545
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

----- Genome Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDrat18
Center clone name: RP11-762624
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator Big-amerham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 14596 bases at least Q40
Consensus quality: 147102 bases at least Q30
Consensus quality: 147971 bases at least Q20
Insert size: 14855; sum-of-contigs
Quality coverage: 10.98x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
13 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

1 44828 contig of 44828 bp in length
44929 82602 contig of 37674 bp in length
82703 101344 contig of 18642 bp in length
101445 116205 contig of 14761 bp in length
116306 123784 contig of 7479 bp in length
123885 130355 contig of 6471 bp in length
130456 133619 contig of 3164 bp in length
133720 137416 contig of 3697 bp in length
137517 141540 contig of 4024 bp in length
141641 144351 contig of 2711 bp in length
144452 146667 contig of 2216 bp in length
146768 148219 contig of 1526 bp in length
148320 149755 contig of 1436 bp in length

* NOTE: This is a 'working draft' sequence. It currently

```


COMMENT

On Nov 16, 2001 this sequence version replaced gi:15383820.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.sngc.stanford.edu
Quality: Phrap Quality >=40 99.8% of Sequence;
Estimated Total Number of Errors is 0.3.

FEATURES

SOURCE

Location/Qualifiers
1. 163803
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-405L7"
BASE COUNT 54797 a 31372 c 29086 g 48548 t
ORIGIN

Query Match

86.9%; Score 551.8; DB 9; Length 163803;

Best Local Similarity 93.2%; Pred. NO. 5e-165;

Matches 577; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

```
QY 1 cccctatcttaacctctctgttaagtttgtctcttcagaatcaaaactgtaaaacta 60
    |||
Db 117348 CCTGTATCTTTAACTTCTCTGTTAGTTTGTCTCTCCAGATTCAGACTGTAAAGCTA 117289

QY 61 caaatgtctcaaatgagacaccagatggagtcacatgaatccaccgttgacc 120
    |||
Db 117288 CAAATAGTTCTTCAAAATGGAACCCAGATGAGTCATGACTAAATCTACCGTGGACCC 117229

QY 121 ctggacggcctgtctagcccatgctcgatgttaatgaatgaagcaccctccgag 180
    |||
Db 117228 CTGGACCGGCGCTGTAGACTGTGCTGTATGACATTTGAATGACCCCTCCGAG 117169

QY 181 gaaatcctcaactgcagacccctactatgcccgaatlcagcggagagcagttagaagcgt 240
    |||
Db 117168 GAAATCTCACTGCACAAACCCCTACTACATTCCAATTCAGTAGAGCAAGTTAGACAGT 117109

QY 241 catcagccaacctcccaacaacagcacttgggttttcctgttgagaaggggagctgaagac 300
    |||
Db 117108 TGTGAGCCCAACCTCCCAACAGTACTGGTTCCTGTTGAGAGGTGAGTGAAGAC 117049

QY 301 aggaactagctgatttctcagcaagaagaatccctaagcctagctgggaaggtagct 360
    |||
Db 117048 AGGACTAGCTGGATTCTTAGCTGACTAGAGATCCCTAAGCCTAGCTGGGAAGGTGACC 116989

QY 361 gcatccacctctaaacatgagggctgtcaactagctcacaccggaccaatcagaagagctc 420
    |||
Db 116988 GCATCCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCCGACCAATCAGAGAGCTC 116929

QY 421 actaaatgtcaatcagcaaaaataggagtbaaagaatatagccaatcatatgtgctg 480
    |||
Db 116928 ACTAAATGCTAATCAGCAAAACAGAGAGTAAAGCAATAGCCAAATCATCTATTGCGCTG 116869

QY 481 agagacagcgaggaggaaggaatcggatataaaccgaagcattcgaccggcaagcg 540
    |||
Db 116868 AGAGCAGACGCGGGAAGACAGAGATTGGATATTAACCCAGGCAATTCAGCCAGCAACAG 116809

QY 541 caaccgcccttgggtccctcccttggatggcgctgtgtttcacctatltcacctct 600
    |||
Db 116808 CAACCCCTTGGGTCCCTCCCTCCCATTTATGGAGAGCTGTGTTTCACTCTATTTCACCTCT 116749

QY 601 ataaatctgcaactgaa 619
    |||
Db 116748 ATTAATCTTGCAACTGCA 116730
```

Search completed: June 20, 2002, 09:43:49
Job time: 10754 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 08:18:17 ; Search time 3276.58 Seconds
(without alignments)
6100.565 Million cell updates/sec

Title: US-09-319-156a-9
Perfect score: 1481
Sequence: 1 atgcgcctccctatcatcac.....gtctacaatgaaccacca 1481

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_estbda:*
2: em_esthum:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	647	43.7	723	9	AUI38405 AUI38405
2	629.4	42.5	631	12	AQ261133 CTRBT-E1-
3	628.6	42.4	702	9	AUI38097 AUI38097
4	609.8	41.2	714	10	BE734284 BE734284
5	536.4	36.2	586	9	BE019603 BE019603
6	535.2	36.1	790	10	BG572445 BG572445
7	390.2	26.3	427	9	AA776439 AA776439
8	389.8	26.3	718	9	AL601847 AL601847
9	371.4	25.1	586	12	AQ776289 AQ776289
10	356.4	24.1	459	10	R68740 R68740
11	356.4	24.1	608	12	AQ725469 AQ725469
12	355.2	24.0	411	10	BG004254 BG004254
13	349.6	23.6	382	10	BI052569 RC5-GN028
14	349.4	23.6	376	10	BI056595 RC5-GN028
15	347.4	23.5	373	10	BI052559 RC5-GN028
16	347.4	23.5	373	10	BI052567 RC5-GN028
17	346.8	23.4	374	10	BI056593 RC5-GN028

C 18	344.2	23.2	373	10	BI052561
C 19	344.2	23.2	373	10	BI052571
C 20	344.2	23.2	373	10	BI052572
C 21	344.2	23.2	373	10	BI056589
C 22	344.2	23.2	373	10	BI052566
C 23	342.6	23.1	373	10	BI052570
C 24	323.2	21.8	517	12	AQ170271
C 25	317.4	21.4	345	10	BI056592
C 26	309.4	20.9	346	10	BI053386
C 27	308.6	20.8	385	12	AQ062249
C 28	308.2	20.8	367	10	BF990707
C 29	304.8	20.6	379	10	D78692
C 30	289.6	19.6	334	10	BI053385
C 31	281.6	19.0	313	10	BI056596
C 32	274.8	18.6	322	10	BF990902
C 33	268.8	18.1	374	10	T93661
C 34	265.4	17.9	315	10	BI055459
C 35	265.4	17.9	315	10	BI056080
C 36	265.4	17.9	331	10	BI055458
C 37	265.4	17.9	331	10	BI056079
C 38	264.8	17.9	300	10	BI055456
C 39	264.8	17.9	300	10	BI056077
C 40	262	17.7	302	10	BG004247
C 41	258	17.4	317	10	BI056590
C 42	249.8	16.9	289	10	BI053219
C 43	249.6	16.9	316	10	BI053220
C 44	240.8	16.3	433	10	N77302
C 45	238.2	16.1	449	10	N58614

ALIGNMENTS

RESULT 1
LOCUS AUI38405 723 bp mRNA linear EST 25-OCT-2000
DEFINITION AUI38405 PLACE1 Homo sapiens CDNA clone PLACE1008489 5', mRNA
sequence.
ACCESSION AUI38405
VERSION AUI38405.1 GI:10999926
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (bases 1 to 723)
Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,
Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and
Isogai, T.
HRI human CDNA project
Unpublished (2000)
CONTACT: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomesehri.co.jp
HRI human CDNA Project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; CDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

FEATURES

source
1..723
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE1008489"
/clone_lib="PLACE1"
/tissue_type="Placenta"
/note="Vector: pMT185FL3"
location/Qualifiers

BASE COUNT 210 a 181 c 130 g 197 t 5 others
ORIGIN

Db 601 AGCAAGAGTCTAGGACAGACTAGTACTGCG 631

RESULT 3
LOCUS AUI38097 702 bp mRNA linear EST 25-OCT-2000
DEFINITION AUI38097 PLACE1 Homo sapiens cDNA clone PLACE1007839 5', mRNA
sequence.
ACCESSION AUI38097
VERSION AUI38097.1 GI:10999618
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 702)
Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3951
Fax: 81-438-52-3952
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
FEATURES
source
1..702
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE1007839"
/clone_lib="PLACE1"
/tissue_type="placenta"
/note="Vector: pME18SFL3"
BASE COUNT 206 a 173 c 131 g 189 t 3 others
ORIGIN

Query Match 42.4%; Score 628.6; DB 9; Length 702;
Best Local Similarity 94.6%; Pred. No. 1.5e-156;
Matches 660; Conservative 0; Mismatches 37; Indels 1; Gaps 1;

Db 633 agaaataaacacacactccgttttagtagtagacactctgtttccaalcctgaaataaccca 692
1 AGAATATAACACACTCTCGTTTGGTAGACACTCTTGTTCACATCTGGAATATAACCA 60
Qy 693 tactcaaacctcacctgtgttaaaattagcaactatagaacacacagctcccaatg 752
Db 61 TACCTCAAACTCACTCGTGTAAATTTAGCAATACATACACAACTCCCAATG 120
Qy 753 catagagtggtaaacacacacacacagaaatagctgtcctaccctcaggaataattttgt 812
Db 121 CATCAGGTGGTAACTCTCCACACAAATAGTCTGCTTACCCTCAGGAATATTTTGT 180
Qy 813 ctgtgtgtaacctcagcctcatcattgttgaatgagctctcagaatcatatgtctctc 872
Db 181 CTGTGTGTAACCTCAGGCTATCGTGTGTAATGGCTCTTCAGAACTATGTGCTCTCTC 240
Qy 873 attcttagtgccctcctatagacacatctacacgaagaattatatacatcatgtctacc 932
Db 241 ATTCTTAATGCCCCCTTAAGACCATCTACACGAAACAATTTATACAGTTAATGTCATATC 300
Qy 933 taagcccaacaaagaagatcaccatctctctcttcttcttctcagcagagagtgctag 992
Db 301 TAAGCCCCGCAACAAAGATACCATCTCTCTTTTGTATAGAGAGAGGAGTCTTAGG 360
Qy 993 cagactagtgactgtgcatgtgcagatcacacaactctactcagttctactacaactatc 1052

Db 361 TGCACTAGTACTGSCATTGGGGGTATCATCAACCTCTACTGAGTCTACTACAACTATC 420
Qy 1053 tcaagaataaatgtgtgacatggaacaggtcacttccctgtgtcaccctgcaagatca 1112
Db 421 TCAGAACTAAATGAGGAGATGGAACGGGTGCGCCACCTCCCTGGTGCACCTTGCAAGATCA 480
Qy 1113 acttaacctccagcagcagtagtctcctcaaatcgaagaagctttagactgtcctaacgc 1172
Db 481 ACTTAACCTCCCTAGCAGCAGTAGTCTTCAAAATCGAAGAGCTTTAGACTGTCTAACCGC 540
Qy 1173 caaagaaggaggaaacctgttatttttagagaagaagcagctgtatatagttaacatc 1232
Db 541 TGAAGAGGAGGGGAACCTGTTTATTTTAGGGGAAGAAATGCTTTATTTATATCAATC 600
Qy 1233 cagaatgtcactgagaagaatgaaagaatccagatcgaaatcagatagtag-agcagagg 1291
Db 601 CGAATCGTCACTGAGAAAGTTAAAGAAATTCAGATGTGAATACACGTANNAAGCANNAG 660
Qy 1292 agcttcaaacacacgaacgctggggcctccctcagccaa 1329
Db 661 AGCTTCGAAACACTGGACCTGGGGCTCTCTAGCCAA 698

RESULT 4
LOCUS BE734284 714 bp mRNA linear EST 15-SEP-2000
DEFINITION BE734284 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3840572 5',
mRNA sequence.
ACCESSION BE734284
VERSION BE734284.1 GI:10148276
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 714)
NIH-MGC <http://mhc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov
Plate: LCM530 row: m column: 21
High quality sequence stop: 712.
FEATURES
source
1..714
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3840572"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 194 a 183 c 137 g 200 t
ORIGIN

Query Match 41.2%; Score 609.8; DB 10; Length 714;
Best Local Similarity 94.3%; Pred. No. 1.5e-151;
Matches 676; Conservative 0; Mismatches 37; Indels 4; Gaps 4;

QY	723	caatctaatgagcacaacacagctcccaatgacaaagcgaagat	782
Db	1	CAATCTACATACACACCAACTCCCAATGCATCAGTGGGTACTCTCCACACAAAT	60
QY	783	agctcgctaccctcaagaaatacttttgcgtggttaacctagcctacatctgttga	842
Db	61	ACTGTGCTTACCCCTCAGGAATATTTTGTGTGTGTACTCAGCCTATTC - GTGTTGAA	119
QY	843	tggctctcaagaatctatgttgtctctcctcatcttctagtggccctctatgacatctaac	902
Db	120	TGGCTCTTAGATCTATGTGCTTCTCTCATCTTAAAGTCCCTATGACATCTACAC	179
QY	903	tgaacaagaattatcaaatcatgctcgtaactaagcccccacaagaagtaccatct	962
Db	180	TGAACAAGATTATACAGTTATGCTATATCTAAGCCCCGCAACMAAAGACTACCATTTCT	239
QY	963	tcccttctgttacagagcagaagtgctcaagcagactaggtaactggtcattgtgcagtacac	1022
Db	240	TCC - TTTGTTATAGAGAGCGAGAGTCTAGTGCACATAGTACTGGCA - TGGCGGTATCAC	297
QY	1023	aacctctactcgtctctactacaactatctcaagaataaattggtgtaactgtgaacagt	1082
Db	298	AACCTTACTCAGTCTTACTACAACTATCTAAGAACTAATGGGACATGGAAGGGCT	357
QY	1083	cactgactccctggtgcaaccttgcaagatcaacttaactccctcagcagcagtagtctctca	1142
Db	358	CGCCGACTCTCCGTGGTACCTTCCAAAGATCACTTAATCTCCTAGCAGCATGTCTCTTA	417
QY	1143	aaatcgaaagagctcttagacttgctaacccgcacaaagaagggggaacctgttactttagg	1202
Db	418	AAATGGAAGAGCTTATGAGACTTGTCTAAACCGCTGAAGAGGGGGAACCTGTTTATTTTAGG	477
QY	1203	agaagaagcgttttttaagttaactcaatccagaattgctcactgtgaaagttaaagaat	1262
Db	478	GGAAGATGCTCTTATATGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGAAAT	537
QY	1263	tcgagaatcgaaatcaaatgtgagcagaagagcttcaaaaacacagaaagcttgggccctct	1322
Db	538	TGAGATGCAATATACAACGTAAGAGCAGAGAGGCTTGGAACACTGGAGCCTGGGGCTCCT	597
QY	1332	cagcaaatgagatgacctgggtctctccctctctttaggaacctctagaagctctaatatgctt	1382
Db	598	CAGCCAATGGAAGCCCTGGATTCCTCCCTCTTCTTAGAGCCTCTAGCAGCTATATATTGCT	657
QY	1383	actcctctcttggaacctgtgactcttaacctcctctttaggtctgtctctccaagat	1439
Db	658	ACTCCTCTTTGGAACCTGTATCTTTAA - CTCTTGTTTACTTTGTCCTTCCAGAAAT	713
RESULT 5			
LOCUS	BE019603	586 bp	mrna linear
DEFINITION	baf4f03.y1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:2907101 5', similar to TR:095244 095244 ENVELOPE PROTEIN; contains Alu repetitive element; , mRNA sequence.		
ACCESSION	BE019603		
VERSION	BE019603.1	GI:8279682	EST.
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 586)		
AUTHORS	NIH-MGC http://mgc.ncl.nih.gov/ .		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1995)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Lung Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNI) DNA Sequencing by: Washington University Genome Sequencing Center		

[illegible]

QY 733 gacacacacgctcccaatgcatcaggtgggtaacacccctcccaacgaatagctgccta
 121 TACACAACCACTCCCAATGCATCAGTGGTAACTCCCTCCACACAAATAGTGTGCTTA
 QY 793 cccctcaggaaatatttttctgctgctgtaacctacatattgttgaatgctcttca
 181 CCTCAGAGAAATATTTTGTCTGTGTGACCTGCTATCGTTGTGAATGGCTCTTCA
 QY 853 gaactcagctctccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc
 241 GATTCATATGCTTCCCTCATTTCTAGTCCCTCATGACATCTACACTGACAAAGAT
 QY 913 ttatacaatcagctgctaccta-gccccacaaacgaagtaacca-ttcttccttctg
 301 TTATACAGTATGTCATATCTAAGCCGCCACAAAGAGTACCCATTCTTCTTTTGG
 QY 971 ttat-cagagcagagatgctcaggacagatagctacgtgctcagctatcacacacctc
 361 TTATAGGAGACGAGGGGTGTGGGTGATGGGTGANTGGCGGTATACACACCTTT
 QY 1030 acccagctc-tactacaactcctcgaagaataatgg 1067
 421 ATTCACTTTTATTACAAATTTCTCAGGANTATATGG 459

RESULT 11 608 bp DNA linear GSS 14-JUL-1999
 LOCUS AO725469 HS.5403.A1.E02.T7A RPCI-11 Human Male BAC Library Homo sapiens
 DEFINITION genomic clone Plate-979 Col-3 Row-1, DNA sequence.
 ACCESSION AO725469
 VERSION AO725469.1 GI:5485138
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 608)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Kellier,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieterdejong.med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
 or from Research Genetics (info@resgen.com). BAC end Web Server:
 http://www.htsc.washington.edu
 Plate: 979 row: 1 column: 3
 Seq primer: 17
 Class: BAC ends
 High quality sequence stop: 608.
 Location/Qualifiers
 1..608
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate-979 Col-3 Row-1"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor
 and partially digested with a combination of EcoRI and

EcoRI Methylase. Size selected DNA was cloned into the
 pBACe3.6 vector at EcoRI sites"
 BASE COUNT 165 a 114 c 145 g 173 t 11 others
 ORIGIN

Query Match 24.1%; Score 356.4; DB 12; Length 608;
 Best Local Similarity 88.3%; Pred. No. 4, 8e-84;
 Matches 406; Conservative 0; Mismatches 52; Indels 2; Gaps 2;

QY 1022 caaccctcctcagctcctcctcctcctcctcctcctcctcctcctcctcctcctc
 608 CAACCTCTACTCAGTTTACT-CAACATATCTCAAGATTAATGGGTGACATGAGATGG 550
 QY 1082 tcaactacccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc
 549 TCACNGACTCCCTGGTCCACTTTCAGATCTTCAACTTCACTTCACTGAGTGTCTTC 490
 QY 1142 aaaaacgaagagctttagactgctaaacgcgaagagggggaacctgttatttttgg
 489 -AAATCGAAGAGCTTTAGACTTCTTAACTGCAAGAGAGAGGAACTGTTTATTAG 431
 QY 1202 gagaagaagctgttattatgtaatacaatccgaattgctcactggaagtaagaaa
 430 TGAAGAATGCTGTATTATACGTTAATCAATCTGAATCGCACCGAANAAGTTAAGAAA 371
 QY 1262 ttccagatcgaatacaatgtagagcagagagctcctcctcctcctcctcctcctc
 370 TTCAAGATGCAATATCAACATAGAGAGAGAGGTTTCANANACACAGACCTGGAGACTCG 311
 QY 1322 tcaagcaatgagatgctcctcctcctcctcctcctcctcctcctcctcctcctcctc
 310 TCAGCCAAATGAGATGCCCTGATCTCCCTCTTCTTACGACTCTACGACTATATCTGT 251
 QY 1382 taccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc
 250 TACTCTCTTTTGAACCTGATCTTATACCTCTTTTAAAGTTGCTCTTCAGAAATTC 191
 QY 1442 aagctgtaaagctacagatgctcctcctcctcctcctcctcctcctcctcctcctc
 190 AACCTGTAAGCTAACAAATGGCTCTTCAATATGAGAGCCCA 151
 Db

RESULT 12 411 bp mRNA linear EST 24-JAN-2001
 LOCUS BG004254 RC6-GN0070-301100-014-C06 GN0070 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BG004254
 ACCESSION BG004254
 VERSION BG004254.1 GI:12445242
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 411)
 Dias Neto,E., Garcia Correa,R., Verjowski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Borlin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPSP/JICR Human Cancer Genome Project. This entry can be seen in the following URL (<http://www.judmgi.org.br/scripts/gethtml2.pl?tbl=RC6&t2=RC6-GN0070-301100-014-C06&t3=2000-11-13&t4=1>)

DEFINITION RC5-GN0282-210201-013-G04 GN0282 Homo sapiens cDNA, mRNA sequence.
ACCESSION BI056595
VERSION BI056595.1 GI:14464125
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 376)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,M. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?rl=RC5&rl2=RC5-GN0282-
210201-013-G04&rl3=2001-02-21&rl4=1)
Seq primer: puc 18 forward
High quality sequence stop: 376.
Location/Qualifiers
1. 376
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="GN0282"
/dev_stage="Adult"
/note="Organ: placenta.normal; Vector: puc18; Site:1: SmaI
; Site:2: SmaI; A mini-library was made by cloning
products derived from ORESTES PCR (U.S. Letters Patent
Application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the pUC 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."
BASE COUNT 90 a 52 c 120 g 114 t
ORIGIN
Query Match 23.6%; Score 349.4; DB 10; Length 376;
Best Local Similarity 95.7%; Pred. No. 3.2e-82;
Matches 359; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 418 catagaccccttagccctctcaagaagactagtctctcaaaactatgaaccctcgct 477
DB 376 CATGACACCTCTAGCCCTCAAGAGAGAGATCTCTCAAACTACATGAAACCCCTCCG 317
QY 478 accatactcgctgtagcctatttaataaccactcactcggtctcagtgctca 537
DB 316 ACCCATACTGCGCTGTAAGCCCTATTTAACCACTCCTGCGCTCATGAGGCTCG 257
QY 538 gcccaaaacccctactactgttgatgctgcctccctgcacttcagggcatatctca 597
DB 256 GCCCAAAACCCCTACTAAGCTTGATATGCGCTCCCTCGAAGTTTCAAGGCATATGTTTCA 197
QY 598 atccctgcttctgaacaatggaacacttcagacagaagaataaacaccacttcgtttca 657
DB 196 ATCCCTGTACTGTAACAATGGAACAACCTTCAGCAGAAATAAACACCACTTCCGTTTAA 137
QY 658 gttagagcctctgttccaaatctggaataaacataaccctcaaacctcaccctgtgtaaa 717
DB 136 GTAGGACCTCTTCTTTCATATCTGGAATTAACCATACCTCAAACTCAGCTGTGTAAT 77

QY 718 tttagaactactatagacaacacgaagctcccaatgcatcagtgaggtaaacacctccaca 777
DB 76 TTTAGCAATFACCTACATACAGCAACCAACTCCCATGATCATGAGTGGAAGTACTCTCCACA 17
QY 778 cgaatagctgcgcta 792
DB 16 CAAATAGTGTGCTTA 2
RESULT 15
BI052559/c 373 bp mRNA linear EST 15-JUN-2001
LOCUS RC5-GN0282-220101-011-A04 GN0282 Homo sapiens cDNA, mRNA sequence.
DEFINITION BI052559
ACCESSION BI052559
VERSION BI052559.1 GI:14460089
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 373)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,M. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?rl=RC5&rl2=RC5-GN0282-
220101-011-A04&rl3=2001-01-22&rl4=1)
Seq primer: puc 18 forward
High quality sequence stop: 373.
Location/Qualifiers
1. 373
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="GN0282"
/dev_stage="Adult"
/note="Organ: placenta.normal; Vector: puc18; Site:1: SmaI
; Site:2: SmaI; A mini-library was made by cloning
products derived from ORESTES PCR (U.S. Letters Patent
Application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the pUC 18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."
BASE COUNT 89 a 52 c 120 g 112 t
ORIGIN
Query Match 23.5%; Score 347.4; DB 10; Length 373;
Best Local Similarity 95.7%; Pred. No. 1.1e-81;
Matches 357; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
QY 418 catagaccccttagccctctcaagaagactagtctctcaaaactatgaaccctcgct 477
DB 373 CATGACACCTCTAGCCCTCAAGAGAGAGATCTCTCAAACTACATGAAACCCCTCCG 314
QY 478 accatactcgctgtagcctatttaataaccactcactcggtctcagtgctca 537

```

Db 313 ACCCATCTCGCCTGTGTAAGCCTATTTAATACACCTTCACCTGGGCTCCATGAGGTCCTCG 254
OY 538 gcccaaaacccctactaaactgttggatgtgacctccccctgcaacttgaaggccatatacttca 597
Db 253 GCCCAAAACCCCTACTACTGTGGATATGCTCCCTGAACTTCAGGCCATATGTTTCA 194
OY 598 atccctgtcctgaacaatggaacaacttcagcaagaataaacaacacacttcglttla 657
Db 193 ATCCCTGTACCTGACGACATGGAACAACACTTCAGCAGAGAATAATACACCCACTTCGGTTTCA 134
OY 658 gtagaacctctgtttccaatcgtgaataaaccatacctcaaacctcactgtgtlaaaa 717
Db 133 GTAGGACCTCTGTTTCCAATCTGGAATAACCCATACCTCAAACTCAGCTGTGTAAAA 74
OY 718 tttaagaatactatagaacaacacagctcccaatgcatcaggtggtaaacactccaca 777
Db 73 TTTAGCAATACTACTACACACCAACTCCCAATGATCAGGTGGGTAACTCCTCCACACA 14
OY 778 cgaatagctgtgcc 790
Db 13 CAAATAGTCTGCC 1

```

Search completed: June 20, 2002, 08:18:31
 Job time: 7357 sec



Db 1168 catgcaatactcatattatgacacagaaaaatgaattaatcctaattatctgaagact 1227

Qy	301	ggagcgacgtgctctgtgactctcttaccatcaacgaatgctgtatgvgggtggaatt	360
Db	1228	ggagttcaacgtctgtgtgacttacttaccacaaactgttatgtctgtatgggggtgagtt	128
Qy	361	caagtcgagcgaaagagaaaaaacaagtataaagaaagcaatctcccaactgtgcccgggagcat	420
Db	1288	caagatcagagaaagaaaaaactgtataaagaagttaatctcccaactcaacccgggtact	134
Qy	421	agcacccctagccctcaaaagactagtctctcaaaactatagaacccctcgtacc	480
Db	1348	ggcacctctagccctcaaaagactagatctctcaaaactatagaacccctcgtacc	140
Qy	481	catctctgcggcgagacccatttaatacacacccctactcgtctccatagtgtctcagcc	540
Db	1408	catctctgcggcgagacccatttaatacacacccctcaactcgtgtctccatagtgtctcagcc	146
Qy	541	caaaacccctactaacgtctgtgagtgtccctccctgcactcagctcaggtacatattcaac	600
Db	1468	caaaacccctactaacgtctgtgatactgctccctcgtgaacttcaggtacatatttttaac	152
Qy	601	ccgtctcccgaaacaatgaaacaactcagcacagaataaacaaccacttcggttttagta	660
Db	1528	ccgtaccccgaaacaatgaaacaactcagcacagaataaacaaccaacttcggttttagta	158
Qy	661	ggacacctgttttccaatctcgtggaataaaccaatacctcaacactcaactcgtgtaaatt	720
Db	1588	ggacacctgttttccaatctcgtggaataaaccaatacctcaacactcaactcgtgtaaatt	164
Qy	721	agcaatactatagacacaacacagctcccaatgtgatcaggggtgtaacactcccaaga	780
Db	1648	agcaatactatatacaacaacaacactcccaatgtcatcaggtgtggttaactctctccacaca	170
Qy	781	atagtctgctacccctcaggaataattttgtctgtgtgataccctcagccatactgttgy	840
Db	1708	atagtctgctacccctcaggaataattttgtctgtgtgataccctcagccatactgttgy	176
Qy	841	aatgtgctcttcagaatctatagtgtctctctctcaattcttagtgccctatgacacatc	900
Db	1768	aatgtgctcttcagaatctatagtgtctctctctcaattcttagtgccctatgacacatc	182
Qy	901	actgaacagaattatatacaatctcgtatcctcaagccccaacaaagaagatgaacct	960
Db	1828	actgaacagaattatatacaatctatgttcatactcaagcccgcaaaaaagatgaacct	188
Qy	961	cttcctttgttalcagagagaggtgtctatggcagactagtactgtgcatctggcagtac	1020
Db	1888	cttcctttgttalcagagagaggtgtctatggcagactagtactgtgcatctggcagtac	194
Qy	1021	acaaacctctactcagttctctctacaaactctccaagaataaagtgtacatgtgaacg	1080
Db	1948	acaaacctctactcagttctctctacaaactctccaagaataaagtgtgaacatgtgaacg	200
Qy	1081	gtcaactgactcccggtgcacctctgccaagaatacaacttaaccccttagcagaatgacct	1140
Db	2008	gtcgcgcaactcccggtgcacctctgccaagaatacaacttaaccccttagcagagtagacct	206
Qy	1141	caaaactcgaagagctcttagactctgtctaacccgcaaaaggggggaacccgttatattta	1200
Db	2068	caaaactcgaagagctcttagactctgtctaacccgcaaaaggggggaacccgttatattta	212
Qy	1201	ggagaaagaacgctgttatctatgttaatcaatccagaatgtgcactgtagaagaagttaaga	1260
Db	2128	ggggaagaatctgtttatcttgtaatcaaatccggaatcgctacatgagaagttaaga	218
Qy	1261	attcgagatcgaataaataatagagcagagagcttcaaaacccgaacgcggggccct	1320
Db	2188	attcgagatcgaataaacaacgtacagcagagagcttcaaaacccgtggagccct	224
Qy	1321	ctcagcaaatgtgacccctgtgtctccctctcttagacccttaagcagctcaaatgt	1380
Db	2248	ctcagcaaatgtgagccctgtgatctccctctcttagacccttagcagctataatgt	230

Oy	1361	ttaactcccttggagcccttatcattaaacccctgtgtaattgtgccttcagaatt	1440
Db	2308	cctaactcctttggagcccctgtatcctttaaccctccctgtgaacttgccttcagaattc	2367
Oy	1441	gaagctgttaaagctacagatagtgcctatacaaatggaaaccca	1491
Db	2368	gaagctgtuaaaactacaatatggagagccccagaatgacatccaa	2408

RESULT 2

US-08-686-878A-48
; Sequence 48, Application US/08686878A

; Patent No. 5708157
; GENERAL INFORMATION:
;

APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John

APPLICANT: LaVallie, Edward
APPLICANT: Racie, Lisa

```

; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT:

```

```

;
; APPLICANT: Evans, Cheryl
; APPLICANT: Spaulding, Vikki
;

```

TITLE OF INVENTION:	SECRETED PROTI
TITLE OF INVENTION:	ENCODING THEM

```

; NUMBER OF SEQUENCES:  /1
; CORRESPONDENCE ADDRESS:
;

```

ADDRESSEE: Genetics Institute,
STREET: 87 CambridgePark Drive

```

; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA

```

COUNTRY: U.S.A.
ZIP: 02140

```

; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
;
; COVERED: 100%

```

```
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

; SOFTWARE:  PatentIn Release #1.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  USC08/6885 8

```

APPLICATION NUMBER: 05/08/0000,8
FILING DATE:
CLASSIFICATION: 536

CLASSIFICATION: 330
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A

NAME: BROWN, SCOLL A.
REGISTRATION NUMBER: 32,724
TELECOMMUNICATION INFORMATION:

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851

```

; TELEFAX: (017) 870-2821
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 542 base pairs
; TYPE: nucleic acid

```

STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: CDNA
;
JIS-08-686-878A-48

6
6
6
6
6
5
3
4
6

Query Match	Score
Best Local Similarity	91.58
Best Global Similarity	91.58
Best Local Similarity	91.58
Best Global Similarity	91.58

Matches 184; Conservative 0; M...

QY 1 atggccctccctatcatactttctctt

Db 342 ATGCCCTCCCTTATCATATTTTCTCTT

QY 61 gcacccctccatgtgtgtacaaccag

Db 402 GCACCCCTCCATGCCGCTGTATGACCAG

QY 121 acgcgcttcctggaatatgtgacccc

Db 462 ATGCAGCGTCCCGGAATATTGATGCCCC

```

Oy 1361 taccctcttgagccctgtatctcttaaccctctgttaagttgtctctccaagatt 1440
      |||||
Db 2308 ctaccctcttgagccctgtatctcttaaccctctgttaagttgtctctccaagatt 2367

Oy 1441 gaagctgtaaagctacagatggtcttacaatggaaccca 1481
      |||||
Db 2368 gaagctgtaaactacaatgagcccaagatgagtcctca 2408

RESULT 2
US-08-686-878A-48
/ Sequence 48, Application US/08686878A
/ Patent No. 5708157
/ GENERAL INFORMATION:
/ APPLICANT: Jacobs, Kenneth
/ APPLICANT: McCoy, John
/ APPLICANT: Lavallie, Edward
/ APPLICANT: Racie, Lisa
/ APPLICANT: Merberg, David
/ APPLICANT: Treacy, Maurice
/ APPLICANT: Evans, Cheryl
/ APPLICANT: Spaulding, Vikki
/ TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
/ NUMBER OF SEQUENCES: 71
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Genetics Institute, Inc.
/ STREET: 87 Cambridgepark Drive
/ CITY: Cambridge
/ STATE: Massachusetts
/ COUNTRY: U.S.A.
/ ZIP: 02140
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentln Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/686, 878A
/ FILING DATE:
/ CLASSIFICATION: 536
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Brown, Scott A.
/ REGISTRATION NUMBER: 32,724
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617) 498-8224
/ TELEFAX: (617) 876-5851
/ INFORMATION FOR SEQ ID NO: 48:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 542 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
US-08-686-878A-48

Query Match 11.7%; Score 173.8; DB 1; Length 542;
Best Local Similarity 91.5%; Pred. No. 1.4e-46;
Matches 184; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Oy 1 atgagccctccattacatactcttctcttactagtctcttaccctctcgctctcaact 60
      |||||
Db 342 atggccctcccttancatattttttcttactgtttcttttaccctcttctcact 401

Oy 61 gaacccctcatctgtctgttacaacacgtagctcccttaccacaagagttctatgaaga 120
      |||||
Db 402 gcacccctcttcacatgcctgtatgacacagtagctcccttaccacaagagttctatgaaga 461

Oy 121 acgagctctctggaatcatatgatgccccatcataatagagttctcaagggaaaccc 180
      |||||
Db 462 atgcagcgtctcccgaaatatttgatgcccccattcgtatagagacttcttctaagggaaccccc 521

```


[illegible]

SULT 5
US-08-691-563C-46
Sequence 46, Application US/08691563C
Publication 6001087

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

ADDRESS: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691,563C

BILLING DATE: 02-AUG-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Berridge, William P.
 REGISTRATION NUMBER: 30,024
 REFERENCE/DOCKET NUMBER: WMB 38588
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-936-6400
 TELEFAX: 703-936-2787
 INFORMATION FOR SEQ ID NO: 46:
 SEQUENCE CHARACTERISTICS
 LENGTH: 1859 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-08-691-563C-46

Query Match	8.9%	Score 131.2	DB 3	Length 1859
Best Local Similarity	57.6%	Pred. No. 2e-32		
Matches 257	Conservative 0	Mismatches 183	Indels 6	Gaps 1

1036 ttctactacaactatctcaagaaataatggtgacatggaacaggtcactgactccctg 1095

Db	1092	TACTACACACACTCTCAAGGATTTCTCAGACAGTTTGCCAGAAATATGATATCATTC	1151
Oy	1096	gtcaacttgcaagatcaacttaacccccctagcagcagtagtccctcaaatcgaaagagct	1155
Db	1152	CTTACTCTACAAATCCCAAATAGACTCTTTGGCAGCAGTACTCTCCAAACCGTCAGAGC	1211
Oy	1156	ttagacttgctaacccgcaaaagagggggaactgtttatttlttggaggaagaaacgcgtc	1215
Db	1212	CTAACACCTCCCTCACTGCTGTGAAAGAGAGACTCTGCACCTTTTAAGGAAAGAGTGTGT	1271
Oy	1216	tattatgttaatcaatccaaactgtgcactgagaaagttaaagaattcgcagatcgaa ta	1275
Db	1272	CTTTACACTAACCGATCAGGATAGTATGATAGCTGCCCGCATTTACAGAAAAGCT	1331
Oy	1276	caatgtagagcagagaagacttcaaaaacacggaacgctggggccctccagccaatygatg	1335
Db	1332	TCTGAATCAGACACACGCTTTCAAAATTC-----CTATACACACCTGTGAGTTGGCA	1385
Oy	1336	ccccgggtctcccccctcttagaacctctagcagcctctaatatgtttactcctcttggaa	1385
Db	1386	ACATGGTTTTTCCCTTTTATCTCATATCTCCATGGCGCCATCTTCTATTTACTCGCCATTTGGG	1445
Oy	1396	ccctgatactttaaccctcctgttaagttgtctctcttcagaatgtgaagcgttaagagta	1455
Db	1446	CCCGTATTTTTTAACCTCTTGTCAAATTTGTTCTTTAGAGATGAGAGCATCAAGCTA	1505
Oy	1456	caga tgc tctacaat tgaagacc ca 1481	
Db	1506	CAGATGGCTTACAAATGAGAACCCCA 1531	

RESULT 6
US-08-007-282B-1
; Sequence 1, Application US/08007282B

Patent No. 3403582
GENERAL INFORMATION:
APPLICANT: NAZERIAN, KEYVAN
APPLICANT: CALVERT, JAY G.
APPLICANT: WITTER, RICHARD L.
APPLICANT: YAMAGIDA, NOBORU
TITLE OF INVENTION: VACCINE COMPRISING FOWLPOX VIRUS
TITLE OF INVENTION: RECOMBINANTS EXPRESSING THE ENVELOPE GLYCOPROTEIN OF AN AVIAN RETICULOENDOTHELIOSIS RETROVIRUS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
STREET: 8110 Gatehouse Road
CITY: Falls Church
STATE: VA

1 COUNTRY: USA
2 ZIP: 22042
3
4 COMPUTER READABLE FORM:
5 MEDIUM TYPE: FLOPPY disk
6 COMPUTER: IBM PC compatible
7 OPERATING SYSTEM: PC-DOS/MS-DOOS
8 SOFTWARE: Patent Release #1.0, Version #1.25
9
10 CURRENT APPLICATION DATA:
11 APPLICATION NUMBER: US/08/007,282B
12 FILING DATE: 19930121
13 CLASSIFICATION: 424
14
15 ATTORNEY/AGENT INFORMATION:
16 NAME: MURPHY JR., GERALD M.
17 REGISTRATION NUMBER: 28,977
18 REFERENCE/DOCKET NUMBER: 1644-104P
19 TELECOMMUNICATION INFORMATION:
20 TELEPHONE: 703-205-8000
21 TELEFAX: 703-205-8050

```

? TELE: 248345
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1704 base pairs
? TYPE: NUCLEIC ACID
?
```

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
US-08-007-282B-1

Query Match 7.9%: Score 116.8; DB 1; Length 1704;
Best Local Similarity 51.5%: Pred. No. 9.2e-28;
Matches 268; Conservative 0; Mismatches 252; Indels 0; Gaps 0;

Qy 932 ctaagccccacaacaagaagataccatctctcttctgtatcacagagagtgctag 991
Db 1178 CTAGCCGTCATAGAGGGCAGTCCAGTTATCCCTCTGTGGCTCTAGGAGTTTCA 1237
Qy 992 gcaagactggtactgagcttgagctatcacacactctactcagttctactaacaact 1051
Db 1238 GGGCTACACTTGTGCTGGAACGGGGCTTGGGCTCTCCGTTCCACTTATCACAAGCTCT 1297
Qy 1052 ctcaagaataaagtgtacatgagacagtcactgactcctgttcaacttgcagaatc 1111
Db 1298 CTATTCATTTGATTTGAGATGTCCAGGCTCTTTCAGGACATCAATGACTACAGAC 1357
Qy 1112 aacttaactccttagcagcagtagtctctcaaatcgaagagctttagacttgcacag 1171
Db 1358 AGATTGATCTCCCTGCTGAGTGTGCTTACAAATAGAGAGGGTTAGACCTATGACTG 1417
Qy 1172 ccaaaagaggggaacctgttatttttagagaagaacgcgtgtattatgttaacat 1231
Db 1418 CCGAACAAGAGGAGATATGTCTCCGACCTCCAGAGAGAGTGTGTTTACGCTACAAGT 1477
Qy 1212 ccaagattgtactgtcagaaagttaaagaattcgagatcgaatacagtagagagagg 1291
Db 1478 GGGTATGTCAGTACAGATGCGAATACTCCAGAGAGACCTTATCGAGAGAAAACTG 1537
Qy 1292 agctcaaaacacgcaagcgtgagcctctcctcagccaatgtagcctgtgtccct 1351
Db 1538 CACTGTACGACAAACCCCTGTGGAGCGGCTTGAAGGGCTTCCATATTGTGACCT 1597
Qy 1352 tcttagagcctctagcagcttaataatgttactcctcttggagccctgtattctaac 1411
Db 1598 TGTTAGGCCCCCTGTTTGGGCTCATATTGTTCCATACCTCGGCCCGGTGATTAAGA 1657
Qy 1412 tccctgttaagttgtctctccagaattgaagctgttaa 1451
Db 1658 CCTGACTCGCATTTACATGACAAATTCAGCGAGTAA 1697

RESULT 7
US-09-078-294-4
Sequence 4, Application US/09078294
Patent No. 6265211
GENERAL INFORMATION:
APPLICANT: Choo, Kong-Hong Andy
APPLICANT: Du Sany, Desiree
APPLICANT: Cancilla, Michael R.
TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
FILE REFERENCE: Davies Col
CURRENT APPLICATION NUMBER: US/09/078, 294
CURRENT FILING DATE: 1998-05-13
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 80246
TYPE: DNA
ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

Query Match 7.8%: Score 114.8; DB 4; Length 80246;
Best Local Similarity 55.7%: Pred. No. 3.8e-26;
Matches 263; Conservative 0; Mismatches 202; Indels 7; Gaps 2;

Qy 1010 ttggcagtatccacaccttactcagttctactacaaactatccaagaataatggtg 1069
Db 58509 ttaacagcgtaacacgtatttaccactactcactacacactctccaagaatttctagaca 58568
Qy 1070 acatggaacaggtcactgactcctctgctacaccttgcaagatcaacttactccttagcag 1129
Db 58569 gtttgcataaagaacgaaatgttcttacttactacaaatcccaatagactcttggcag 58628
Qy 1130 cagtagtccctcaaaaatcgaagagctttagacttgcacgcgcaaaagagggggaacct 1189
Db 58629 cagtagctctccaaaacgcgtgagcctagactcttacttctgtgagaaggaagattct 58688
Qy 1190 gttattttaagagaagaacctgttattatgttaatcaaccagaattgtcactgtga 1249
Db 58689 gac-ttcttaagggttagagtggttcttactaacaccacagcaggaaataatagata 58747
Qy 1250 aagttaaagaattcgaatcgaatacaatgttagcagagaggttcaaacaccgaaac 1309
Db 58748 ccaccagtggtttagaggaagagcttctgaatcagacaatgaccttcaacttata 58807
Qy 1310 gctggggcctctcagcaatgagtgccctgggttccctctttagagactctagcag 1369
Db 58808 ccaa-----cctctggagtggtggcagacatggtctcccttcttagtcttgagcag 58861
Qy 1370 ctctaattgttactcctcttggagacctgtatcttacttactcctgtttaggttgc 1429
Db 58862 ccacttgctaacatgctgacatttggccctgtatcttacttacttcttggcgaatttgc 58921
Qy 1430 ctccagaattgaagctgttaaagctacagatgtcttcaaatggaaccoca 1481
Db 58922 cctctagatcgagggccatcaagctacagatgatcttcaaatgttaaccoca 58973

RESULT 8
US-09-011-745-1
Sequence 1, Application US/09011745
Patent No. 6165715
GENERAL INFORMATION:
APPLICANT: Collins, Mary KL
APPLICANT: Weiss, Robin A
APPLICANT: Takeuchi, Yasuhiko
APPLICANT: Cosset, Francois-Loic
TITLE OF INVENTION: Expression systems
FILE REFERENCE: 09/011,745
CURRENT APPLICATION NUMBER: US/09/011,745
CURRENT FILING DATE: 1998-06-22
EARLIER APPLICATION NUMBER: PCT/GB96/02061
EARLIER FILING DATE: 1996-08-23
EARLIER APPLICATION NUMBER: GB9517263.1
EARLIER FILING DATE: 1995-08-23
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 2518
TYPE: DNA
ORGANISM: RD114
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)
OTHER INFORMATION: n is any nucleotide
US-09-011-745-1

Query Match 7.7%: Score 114.2; DB 4; Length 2518;
Best Local Similarity 55.0%: Pred. No. 8e-27;
Matches 224; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

Qy 1045 aaacatctccaagaataatggtgtacatgagacaggtctacgactcctgttcaacttg 1104
Db 1834 aaattatcccatcagttatcatctgtgtccatgcttctatccggttaccataaagattta 1893
Qy 1105 caagatcaacttaactcctctagcagcagtagtcttcaaatcgaagagctttagacttg 1164

```
Db 1894 caagaccagtgactcgttaagctgaagtagtctcccaaatagaggagcggacccta 1953
      ||||| | | ||| ||| ||||| | ||||| | | | ||| |
Qy 1165 ctaccgcaaaagaggaggaacctgttatttttagaagaagacgttatatgtt 1224
      ||||| | | ||||| | | ||||| | | ||||| | | ||||| |
Db 1954 ctacggcgaagaagaaggaattttagcctcaagaagaagcgtttttatgtc 2013
      ||||| | | ||||| | | ||||| | | ||||| | | ||||| |
Qy 1225 aatcaatcagaatttgcctgagaagaattgaagaattcgatcgatatacattaga 1284
      || | | ||||| | | | | | | | | | | | | | | | |
Db 2014 aacaagtcaggaatttctgagaacaataaagaaccctcaagaagaattatacaaaagc 2073
      || | | ||||| | | | | | | | | | | | | | | | |
Qy 1285 gacgaggaagcttcaaaaacccgaagcctgggctcctcagccaatgagtccttggt 1344
      || | | ||||| | | | | | | | | | | | | | | | |
Db 2074 agggaaagcctggcaaccacccctctcggaccggtctgcaaggctctctccgtaacctc 2133
      || | | ||||| | | | | | | | | | | | | | | | |
Qy 1345 ctccctctcttagacctctagcagctcaatatgttactcctcttgaccctgtatc 1404
      || | | ||||| | | | | | | | | | | | | | | | |
Db 2134 ctaccctctctggaccctactcaaccctcctactcaataacccttgggccatgctgt 2193
      || | | ||||| | | | | | | | | | | | | | | | |
Db 1405 tttaacctctgttaagttgtctctctccagaattgaagctgttaa 1451
      || | | ||||| | | | | | | | | | | | | | | | |
Db 2194 ttacgctcctcatggccttcaatagtatgaccttaatgctgttaca 2240
      || | | ||||| | | | | | | | | | | | | | | | |
```

```
RESULT 9
US-09-011-745-8
; Sequence 8, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Weiss, Robin A.
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A.
; APPLICANT: Takeuchi, Yasuhiko
; APPLICANT: Cosset, Francois-Lolc
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; CURRENT FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 5865
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
; OTHER INFORMATION: construct
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3611)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3612)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3613)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3614)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3799)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3800)
```

```
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3801)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3802)
; OTHER INFORMATION: n is any nucleotide
US-09-011-745-8
```

```
Query Match 7.7%; Score 114.2; DB 4; Length 5865;
Best Local Similarity 55.0%; Pred. No. 1.3e-26;
Matches 224; Conservative 0; Mismatches 183; Indels 0; Gaps 0;
```

```
Qy 1045 aaatattccagaataatagtgtgacatggaacaaagctacgacccttgcttaacctg 1104
      || ||||| | | ||||| | | ||||| | | ||||| | | ||||| |
Db 2692 aaattatccatcagttatattctgagtcgaagcttaccggtacatacaagaattta 2751
      || ||||| | | ||||| | | ||||| | | ||||| | | ||||| |
Qy 1105 caagatcaacttaacctccttagcagcagtagtccctcaaatcgaaagctttaagctg 1164
      ||||| | | ||||| | | ||||| | | ||||| | | ||||| |
Db 2752 caagaccagtgtagctcgttagctgaagtagtcttccaaatagagggagctggacccta 2811
      ||||| | | ||||| | | ||||| | | ||||| | | ||||| |
Qy 1165 ctaccgcaaaagaggagggaacctgttatttttaggagaagaacgctgtattatgtt 1224
      ||||| | | ||||| | | ||||| | | ||||| | | ||||| |
Db 2812 ctacggcagaacaaggaaggaatttcttagccttcaagaanaatgctgttttatgtc 2871
      || ||||| | | ||||| | | ||||| | | ||||| | | ||||| |
Qy 1225 aatcaatcagaatttgcctcgtgagaagttaaagaattcgatcgatgaatatactaga 1284
      || | | ||||| | | | | | | | | | | | | | | | |
Db 2872 aacaagtcaggaatttctgagaacaataaagaacccctcaagaagaattatacaaaaagc 2931
      || | | ||||| | | | | | | | | | | | | | | | |
Qy 1285 gacgaggaagcttcaaaaacccgaagccttgggcctcctcagccaatgagtgcccttggt 1344
      || | | ||||| | | | | | | | | | | | | | | | |
Db 2932 agggaaagccttggaaccaaccctctcggaccggctcgcaaggctcttcttcgtaacctc 2991
      || | | ||||| | | | | | | | | | | | | | | | |
Qy 1345 ctccctctcttaggacctctagcagctcaatatgttactcctcttgaccctgtatc 1404
      || | | ||||| | | | | | | | | | | | | | | | |
Db 2992 ctaccctctctggaccctcctactcaaccctcctactcaataacccttggccatgctgt 3051
      || | | ||||| | | | | | | | | | | | | | | | |
Qy 1405 tttaacctctgttaagttgtctctcctcagaattgaagcgttaa 1451
      || | | ||||| | | | | | | | | | | | | | | | |
Db 3052 ttacgctcctcatggccttcaatagtatgaacttaatgctgttaca 3098
      || | | ||||| | | | | | | | | | | | | | | | |
```

```
RESULT 10
US-09-078-294-3
; Sequence 3, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desirée
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies COL
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 80595
; TYPE: DNA
; ORGANISM: Nucleotide sequence of HC-contlg
US-09-078-294-3
```

```
Query Match 6.1%; Score 90.8; DB 4; Length 80595;
Best Local Similarity 57.5%; Pred. No. 2.4e-18;
Matches 206; Conservative 0; Mismatches 142; Indels 10; Gaps 2;
Qy 1124 tagcagcagtagtcttcaaatcgaaagcctttagactgttcaaccgcaagaagagg 1183
      || ||||| | | ||||| | | ||||| | | ||||| | | ||||| |
Db 5885 tggcagcagtagtcttcaaatcgaaagcctttagactgttcaaccgcaagaagagg-- 58916
      || ||||| | | ||||| | | ||||| | | ||||| | | ||||| |
```

QY	1184	gaacctgttatttttttaggaagaagcctgttatatagttaatcaatccagaattgtca	1243
Db	58917	--attctgcacctctcttagggtagagtgcttttttatactaaccaagctcaggatagat	58974
QY	1244	ctggaagaagttaaaagaattcgagatcgacatacaatgtfagagcagagagcttcacaa	1303
Db	58975	gagataccaccacagtggtttacagsgaaaagcctctgtaatcagaataagccttcaaat	59034
QY	1304	ccgaacgcttgggctcctctcagccaatgtagctgcttggtctccctctcttagaacct	1363
Db	59035	c-----ttataccaacctctggagtggttgggagaaatggctctccctctcaggactg	59088
QY	1364	tacgacgtctaaatatgttactcctctctcttggaccctgatactcttaacctcttgtaagt	1423
Db	59089	tgacagcatcttgtaatagtcgaatttgggcctctgatttttaacctcttggccaat	59148
QY	1424	ttgtctcttcagaaattgaagctgtaaaagctacagatgctttcaaatggaacca	1481
Db	59149	ttgttctctcaggatcgaggccaaagcttcacaaatgatactcttcaaatgtgaacca	59206

```

RESULT 11
US-09-011-745-9
Sequence 9, Application US/09011745
Patent No. 6165715
GENERAL INFORMATION:
APPLICANT: Collins, Mary KL
APPLICANT: Weiss, Robin A
APPLICANT: Takeuchi, Yasuhiro
APPLICANT: Cossel, Francois-Loic
TITLE OF INVENTION: Expression systems
FILE REFERENCE: 09/011,745
CURRENT APPLICATION NUMBER: US/09/011,745
EARLIER FILING DATE: 1998-06-22
EARLIER APPLICATION NUMBER: PCT/GB96/02061
EARLIER FILING DATE: 1996-08-23
EARLIER APPLICATION NUMBER: GB9517263.1
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 9
LENGTH: 3925
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Portion of
FEATURE:
OTHER INFORMATION: construct
NAME/KEY: misc_feature
LOCATION: (3910)
OTHER INFORMATION: n is any nucleotide
FEATURE:
NAME/KEY: misc_feature
LOCATION: (3911)
OTHER INFORMATION: n is any nucleotide
FEATURE:
NAME/KEY: misc_feature
LOCATION: (3912)
OTHER INFORMATION: n is any nucleotide
FEATURE:
NAME/KEY: misc_feature
LOCATION: (3913)
OTHER INFORMATION: n is any nucleotide
US-09-011-745-9

```

	Query Match	4.7%	Score 70	DB 4	Length 3925
	Best Local Similarity	49.3%	Pred. No. 2.4e-12		
	Matches 224	Conservative	0	Mismatches 215	Indels 15
					Gaps 1
QY	1014	cagatcaaaccttactcaggttcttactacaactatctcaagaataatggtgacat	1073		

Db	2948	caatgcacctaatcaaaaaccagcagattttagcgcgcttcacgccgctatccagacgaacct	3007
Qy	1074	ggaacagctactgactctcccttgg tcaacctgtgcaga tcaacttaactcccttagacagcagt	1133
Db	3008	caacgaagctcgaaaaatcatcttaccacttagaaaagtcatcgactcgtttgtctgaagt	3067
Qy	1134	agctcccaaaaatcgaagagctttagactgtcctaaaccgcgaagaagggtggaacctgttt	1193
Db	3068	agtcctacagaacccgaagagagccttagattgtccttcccaaaaggaggaggtctctgcgc	3127
Qy	1194	atttttagaggaagacgcgtgtattatgtttaacaatccagaattgtcactgagaaagt	1253
Db	3128	agccctaaagaagaatgtgtttattatgcagacacacaggagactcagtgagagaaagat	3187
Qy	1254	taaaagaattcgcagatcgcgaatacaatgttagagcagagagagcttcaaaaaccgcgaagcgtg	1313
Db	3188	ggccaaactaaggaagaaggttatcatagagacaaaactatttggatcgagcgaagtg	3247
Qy	1314	gggcctcct-----cagccaatgagatgcctcgtgggttccctcttagg	1358
Db	3248	gttcgaaggcgatttaatatagatcccccgggtttaccaccttaatccacatcatgg	3307
Qy	1359	accctcgcagccttaatatgttactcctctttggaccttatcttaacctctgt	1418
Db	3308	accctaatagctactctactgactcttactcctttggacccttgcacttccaaatcgtatgt	3367
Qy	1419	taagttgtctctccgaatttgaagctgaaag	1452
Db	3368	tcaattctgtaaaagcagagctcagtaatgccag	3401

```

1 RESULT 12
2 US-09-011-745-5
3 Sequence 5, Application US/09011745
4 Patent No. 6165715
5 GENERAL INFORMATION:
6 APPLICANT: Collins, Mary KL
7 APPLICANT: Weiss, Robin A
8 APPLICANT: Takeuchi, Yasuhiro
9 APPLICANT: Cosset, Francois-Lolc
10 TITLE OF INVENTION: Expression Systems
11 FILE REFERENCE: 09/011,745
12 CURRENT APPLICATION NUMBER: US/09/011,745
13 CURRENT FILING DATE: 1998-06-22
14 EARLIER APPLICATION NUMBER: PCT/GB96/02061
15 EARLIER FILING DATE: 1996-08-23
16 EARLIER APPLICATION NUMBER: GB9517263.1
17 EARLIER FILING DATE: 1995-08-23
18 NUMBER OF SEQ ID NOS: 29
19 SOFTWARE: PatentIn Ver. 2.0
20 SEQ ID NO 5
21 LENGTH: 6028
22 TYPE: DNA
23 ORGANISM: Artificial Sequence
24 FEATURE:
25 OTHER INFORMATION: Description of Artificial Sequence: Portion of
26 FEATURE:
27 NAME/KEY: misc_feature
28 LOCATION: (3774)
29 OTHER INFORMATION: n is any nucleotide
30 FEATURE:
31 NAME/KEY: misc_feature
32 LOCATION: (3775)
33 OTHER INFORMATION: n is any nucleotide
34 FEATURE:
35 NAME/KEY: misc_feature
36 LOCATION: (3776)
37 OTHER INFORMATION: n is any nucleotide
38 FEATURE:
39 NAME/KEY: misc_feature
40 LOCATION: (3777)
41 OTHER INFORMATION: n is any nucleotide

```

```

: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (3962)
: OTHER INFORMATION: n is any nucleotide
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (3963)
: OTHER INFORMATION: n is any nucleotide
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (3964)
: OTHER INFORMATION: n is any nucleotide
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (3965)
: OTHER INFORMATION: n is any nucleotide
US-09-011-745-5

```

```

Query Match 4.6%; Score 68.4; DB 4; Length 6028;
Best Local Similarity 49.5%; Pred. No. 1e-11;
Matches 217; Conservative 0; Mismatches 206; Indels 15; Gaps 1;

```

```

QY 1030 actcagttctactacaactatctcaagaataatggtgacatggaacaggtcactgac 1089
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2828 acccagcagttgagcagcttcacgctccagacagaccctcaacgaagtcgaaaag 2887
QY 1090 tccctgtagcacttgcagaatcaacttaactccctagcagcagtagtccctcaaatcga 1149
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2888 tcaatcacaactgaagaagcagctgactcgtgtctgaaagtgtcctacagaacgcgc 2947
QY 1150 agagctttagacttgctaacgcgcgaagaagggggaacctgttcttcttaaggagaaga 1209
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2948 agagagcttagattgttcttcttaaggaggaggtctctgagccctcaagaagaaga 3007
QY 1210 cgcgtgtatagttaataatccagaattgtcactgagaagaagttcaagaattcgagat 1269
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3008 tctgtgttcttaagcagacacacaggggagcagtgagagacagatggccaattaaaga 3067
QY 1270 cgaatacaatgttagcagcagaaggagcttcaaaacacgcagcgtgggacctcct----- 1322
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3068 aggcctaaatcagagacaaaactatcttgagacagcagaagtggtctcgaaggcgtgtt 3127
QY 1323 -----cagccaatgagatgcctcgtggtctccctctcttgagacctttagcagctcta 1374
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3128 aataagatcccccgtgttctcaccactaattccaccatcatgagaccttaatagtactc 3187
QY 1375 atattgttactcctcttgagacctgatatctttaacctctcttgaaagtttctcttcc 1434
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3188 ttaactgacttctactcttgagaccttgcaattcctaactgacttagtccaattgttaagaag 3247
QY 1435 agaattgaagctgtaaag 1452
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3248 aggattctcagtagtccag 3265

```

```

RESULT 13
US-08-766-528-1
: Sequence 1, Application US/08766528
: Patent No. 6190861
: GENERAL INFORMATION:
: APPLICANT: Jay A. Fishman
: TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
: NUMBER OF SEQUENCES: 74
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD, LLP
: STREET: 60 State Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109-1875
: COMPUTER READABLE FORM:

```

```

: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/766,528
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/572,645
: FILING DATE: 14-DEC-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Louis Myers
: REGISTRATION NUMBER: 35,965
: REFERENCE/DOCKET NUMBER: MGP-038CP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617)227-7400
: TELEFAX: (617)227-5941
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 8060 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
US-08-766-528-1

```

```

Query Match 4.6%; Score 68.4; DB 4; Length 8060;
Best Local Similarity 53.3%; Pred. No. 1.2e-11;
Matches 144; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```

```

QY 1045 aaactatctcaagaataatgtgacatgagacaggtcactgacctcctggtcaccttg 1104
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1523 AACCTACATCAAGTAATGTATACAGAGATGTCGAAGCCCTGAAATAATCTGTACGTAACTG 1582
QY 1105 caagatcaacttaactcctctagcagcagtagtctccttcaaatcgaagaagctttagactg 1164
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1583 GAGGAATCCCTTAACCTCTTACTGAGAGTAGTCCTACAGAAATGAGAGAGGTTAATTGA 1642
QY 1165 ctaaccgcgaagaaggagggaacctgttatcttcttaaggagaagaacgctgttatctgt 1224
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1643 TTAATTTCTTAAGAAAGAGAGATTAATGTATAGCTTGAAAGAGAGATGCTTTTATGTG 1702
QY 1225 aatcaatccagaattgtcactgagaagaagttaaagaattcgsatcgatatacaatgtga 1284
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1703 GATCAATTCAGGGCCCATCGACATGCAATGAACTTAAGAAAGTTGAGAAAGTTGAGAAAGCGT 1762
QY 1285 gcagaggagcttcaaaaacccgaacgctgg 1314
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1763 CGAAGGAAAGAAAGAACTACTCAAGGGTGG 1792

```

```

RESULT 14
US-08-716-351A-5
: Sequence 5, Application US/08716351A
: Patent No. 6033905
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: Gbbon Ape Leukemia Virus-Based
: NUMBER OF SEQUENCES: 5
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/716,351A
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/03784
: FILING DATE: 06-APR-1994

```

```
ATTORNEY/AGENT INFORMATION:
: NAME: Bastian, Kevin L.
: REGISTRATION NUMBER: 34,774
: REFERENCE/DOCKET NUMBER: 15280-128-1PC
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 543-9600
: TELEFAX: (415) 543-5043
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10970 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: misc feature
: LOCATION: 1..10970
: OTHER INFORMATION: /standard_name="p537 retroviral"
: OTHER INFORMATION: vector"
US-08-716-351A-5

Query Match 4.68; Score 68; DB 3; Length 10970;
Best Local Similarity 56.18; Pred. No. 1.9e-11;
Matches 128; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 1087 gaccctgtgacacccgcaagatcgaacttaaccccttagagcagatgctcctcaaat 1146
DB 6659 GACGCATACAGCAAGCTAGAGGACTCAGCTGCTCCCTATGTGAGTAGTACTCCAAAT 6918
QY 1147 cgaagccttagaactgtcaccgccaagaagggggaacctgttatttttagagaa 1206
DB 6919 AGGAGAGGCGCTTGACTACTTCTCTTAAGAAGAGGCGCTCGCGCCCTTAAGAA 6978
QY 1207 gaagcgtgtattatgttaacatccagaattgcactgagaagttaagaattcga 1266
DB 6979 GAGGCGCTTTTATGTAGACCACTCAGTGCAGTACGAGACTCCATGATAAACTTMA 7038
QY 1267 gatcaatacatgtagacagagagagcttcaaaacccggaagcttg 1314
DB 7039 GAAAGACTAGTAAAGACAGTGTAGAGCCCAAGAAACCAAACTGG 7086

RESULT 15
US-08-258-420-8
: Sequence 8, Application US/08258420
: Patient No. 5710037
: GENERAL INFORMATION:
: APPLICANT: Nienhuis, Arthur W.
: APPLICANT: Vanin, Ello F.
: TITLE OF INVENTION: No. 5710037el Retroviral Envelope and LTR and Retroviral Vecto
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Carrella, Byrne, Baln, Gilfillan, Cecchi, Stewart & Olstein
: STREET: 6 Becker Farm Road
: CITY: Roseland
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07068
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch diskette
: COMPUTER: IBM PS/2
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: Wordperfect 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/258.420
: FILING DATE: 10-JUN-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Olstein, Elliot M.
```

```
REGISTRATION NUMBER: 24,025
: REFERENCE/DOCKET NUMBER: 271010-208
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201-994-1700
: TELEFAX: 201-994-1744
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1911 bases
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: genomic DNA
: FEATURE:
: NAME/KEY: retroviral envelope sequence
US-08-258-420-8

Query Match 4.58; Score 67; DB 1; Length 1911;
Best Local Similarity 48.68; Pred. No. 1.5e-11;
Matches 227; Conservative 0; Mismatches 225; Indels 15; Gaps 1;

QY 1030 actaagtctactcaactatcctaagaataatgtgacatggaacagtgactgac 1089
DB 1417 ACGCAGCAGTTTGACACCTTCATGCCGCTATCCAGACGACCTCAACGAAGTCGAAAG 1476
QY 1090 tcacctgtgacacctgcaagatcaacttaactcccttagagcagatgactcctcaaatga 1149
DB 1477 TCAATTACCAACCTAGAAATACTACACTGCTGTTGTGAGTAGTCTACAGAACCGC 1536
QY 1150 agaacttagaactgtcaccgccaagaagggggaacctgttatttttagagaaga 1209
DB 1537 AGAGGCCCTAGATTGTCTATCTCTTAAGSAGSAGGTCTCTCGCAGCCCTTAAGAAAGAA 1596
QY 1210 cgtgttattatgttaacatccagaattgctcactgagaagttaagaattcagat 1269
DB 1597 TGTGTTTATGTGACAGACACGCGGCTAGTAGAGACACATGCGCCAAATTAAGAGAA 1656
QY 1270 cgaatacatgtagacagagagagcttcaaaacccggaagcttggtctcct----- 1332
DB 1657 AGCCTTAATCAGAGCAAAACCTATTGTGACAGGCCCAAGAGATGTTGGAAGGCTGTTT 1716
QY 1323 -----cagccaatggatgcccttgggttctccctcttagagacctctagcagctcta 1374
DB 1717 AATGATCCCGCTTGTTTACCACCTTAATCTCCACATCATGAGGACCTTAATAGTACTC 1776
QY 1375 atattgtacctcctcttggagccctgtatctttaacctccttgaagttgtctctcc 1434
DB 1777 TTACTGATCTTACTCTTTGGACCTTGCAATCTCAATCGATTAGTCCAATTTGTTAAGAC 1836
QY 1435 agaattgagctgtaaagctacagatggtcttacaagaatgagaccoca 1481
DB 1837 AGGATATCAGTGTGTCAGGCTGTAGTTTGACTCAACATATTCACCA 1883
```

Search completed: June 20, 2002, 09:38:01
Job time: 8391 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 09:42:19 ; Search time 443.18 Seconds

(Without alignments)
5737.511 Million cell updates/sec

Title: US-09-319-156a-9

Perfect score: 1481

Sequence: 1 atgcccctcctatcatcac.....gtctacaagaatgaaccacca 1481

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Optimal number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: N_Geneseq_032802.*
2: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
25: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1481	100.0	1481	19	AAV43217
2	1481	100.0	1481	20	AAV43217
3	1481	100.0	1481	21	AAV43217
4	1461.8	98.7	2030	21	AAV43217
5	1335.4	90.2	2946	20	AAV43217
6	1335.4	90.2	2946	21	AAV43217
7	1333.8	90.1	1617	22	AAH20070
8	1330.6	89.8	2781	22	AAV55630
9	1324.2	89.4	7582	21	AAV59215

10	1322.6	89.3	7582	20	AAV25665	Complete human end
11	1308.2	88.3	2782	22	AAH20069	HERV-W envelope pr
12	1306.6	88.2	2782	20	AAV25661	Human endogenous r
13	1306.6	88.2	2782	21	AAV59211	5' non coding, 3'
14	1311.6	76.4	1894	22	ABV45822	Human breast cell
15	1311.6	76.4	1894	22	ABV45837	Human foetal liver
16	1311.6	76.4	1894	22	ABV45837	Probe #4444 for ge
17	1311.6	76.4	1894	22	ABV45837	Human brain expres
18	1311.6	76.4	1894	22	AAV30018	Human bone marrow
19	1311.6	76.4	1894	22	AAV14608	Probe #4541 for ge
20	1311.6	76.4	1894	22	AAV15980	Probe #4666 used t
21	1311.6	76.4	1894	22	AAV104422	Probe #4413 used t
22	1311.6	76.4	1894	22	AAV25659	Human endogenous r
23	1311.6	76.4	1894	21	AAV59209	5' non coding, 3'
24	1304.8	70.8	6394	23	AAV584210	DNA encoding novel
25	716.6	48.4	792	22	ABV45840	Human breast cell
26	716.6	48.4	792	22	ABV45840	Human foetal liver
27	716.6	48.4	792	22	ABV45840	Probe #14338 for g
28	716.6	48.4	792	22	AAV17251	Human brain expres
29	716.6	48.4	792	22	AAV43040	Human bone marrow
30	716.6	48.4	792	22	AAV23803	Probe #13736 for g
31	716.6	48.4	792	22	AAV19115	Probe #17801 used
32	716.6	48.4	792	22	AAV109412	DNA encoding novel
33	644.6	43.5	902	23	AAV1722	DNA encoding novel
34	644.6	43.5	1039	23	AAV584206	DNA encoding novel
35	644.6	43.5	1341	23	AAV1719	DNA encoding novel
36	644.6	43.5	1341	23	AAV584194	DNA encoding novel
37	582.8	39.4	591	20	AAV25685	Human endogenous r
38	582.8	39.4	591	21	AAV59204	Probe Penv-C15 use
39	564.4	38.1	1088	23	AAV59269	DNA encoding novel
40	564.4	38.1	1174	23	AAV52498	DNA encoding novel
41	564.4	38.1	1929	23	AAV1718	DNA encoding novel
42	564.4	38.1	1929	23	AAV57466	DNA encoding novel
43	564.4	38.1	1929	23	AAV52493	DNA encoding novel
44	564.4	38.1	1929	23	AAV52866	DNA encoding novel
45	516.6	34.9	1136	20	AAV25660	Human endogenous r

ALIGNMENTS

RESULT 1	
AAV43217	standard; CDNA: 1481 BP.
ID AAV43217	
AC AAV43217	
DT 29-DEC-1998	(first entry)
XX	
DE	Multiple sclerosis associated retrovirus fragment 5.
KW	Multiple sclerosis associated retrovirus; MSRV; MS; pol gene;
KW	gag gene; env gene; Rheumatoid arthritis-associated virus; ss.
XX	
OS	Multiple sclerosis associated retrovirus.
XX	
FT	Key
FT	location/qualifiers
FT	1..1479
FT	/*tag- a
FT	/product- "Encodes protein AAV71068"
FT	/transl_except- (pos:115-117, appears to code for a
FT	stop codon)
FT	/note- "CDS does not contain a stop codon"
XX	
XX	W09823755-A1.
XX	
XX	04-JUN-1998.
XX	
XX	26-NOV-1997; 97WO-IB01482.
XX	
XX	26-NOV-1996; 96US-0756429.
XX	
XX	(INMR) BIO MERIEUX.

XX Bedin F, Beseme F, Jolivet-Reynaud C, Komurtan-Pradel F,
PI Mandrand B, Paranhos-Baccala G, Perron H;
XX WPI: 1998-322732/28.
XX P-PSDB: AAM71068.
DR
XX
XX New nucleic acid from retroviruses - useful for diagnosis,
PT prevention and treatment of, e.g. multiple sclerosis
PT
XX
PS Disclosure: Pages 184-185; 286pp; English.

OY	661	ggaacctctgttctccaalctcggaaataaacccatacctcaaacctcaccgtgtgtaaaatt	720
Db	661	ggaacctctgttctccaalctcggaaataaacccatacctcaaacctcaccgtgtgtaaaatt	720
OY	721	agcaatactatagagacacacacagctcccaatgcatacgtggtgttaaacctcccaacga	780
Db	721	agcaatactatagagacacacacagctcccaatgcatacgtggtgttaaacctcccaacga	780
OY	781	atagactcgcacacccctcaggaataatttttttcgtgtgtaccccaagccttaatttttg	840
Db	781	atagactcgcgcacacccctcaggaataatttttttcgtgtgtgtaccccaagccttaatttttg	840
OY	841	aatggctcttcagaatactatgtgtcttcctcctcatctcttagtgcccccattagacacatac	900
Db	841	aatggctcttcagaatactatgtgtcttcctcctcatctcttagtgcccccattagacacatac	900
OY	901	actgacaagaattatatacaatcattgtcttaactaaagcccaacaacaaagatattaccatt	960
Db	901	actgacaagaattatatacaatcattgtcttaactaaagcccaacaacaaagatattaccatt	960
OY	961	cttccttttgtatcatcagagcagagatgctcaggcagactaggtaactctggcatctggcagtatc	1020
Db	961	cttccttttgtatcatcagagcagagatgctcaggcagactaggtaactctggcatctggcagtatc	1020

Query Match	100.0%	Score 1481;	DB 19;	Length 1481;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1481; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0

Db 1021 acaactctactcagttctactacaactatctcagaataaattgtagacatggaacag 108

OY	1	atgagccccccttataaactttctctcttaccgtgtctcttaccctcttcgctctact	60
Db	1	atggcccccccttaacaaacttctctcttaccgtgtctcttaccctcttcgctctact	60
OY	61	gaacccccctcaatgcgtgcgtgtacacaagaatagctccccccttaaccagaattctctatgaaga	120
Db	61	gaacccccctccatctgtgcgtgtacacaagaatagctccccccttaaccagaattctctatgaaga	120
OY	121	acggcgctctcttggaataattgatagccccaacataataggatttatcttaagggaatacttc	180
Db	121	acggcgctctcttggaataatattgatagcccccaacataataggatttatcttaagggaatacttc	180
OY	181	acctcaacttgcacccacaacccaatgatcccccgcaactgctatactctgcgaactctttgcatg	240
Db	181	accttcaacttgcacccaacccaacatgctcccgcaactgctatactctgcgaactctttgcatg	240
OY	241	catctcaaatatctctatttttgagacgggaaaaatgatactctgatttctcttgagagact	300
Db	241	catctcaaatatctctatttttgagacgggaaaaatgatactctgatttctcttgagagagact	300
OY	301	gagagccaactgtctcgtgtgacttacttcaaccataccagatagtcctgaatgggggttggaatt	360
Db	301	gagagccaactgtctcgtgtgacttacttcaaccataccagatagtcctgaatgggggttggaatt	360
OY	361	caaaagtctcagagaaagaaaaacaagtataaggaaagaaatctcccaactgacccggggagcat	420
Db	361	caaaagtctcagagaaagaaaaacaagtataaggaaagaaatctcccaactgacccggggagcat	420
OY	421	agaccccccttaagccccctacaagaagactagtctctccaagaactacatgaaacccctcgtatcc	480
Db	421	agaccccccttaagccccctacaagaagactagtctctccaagaactacatgaaacccctcgtatcc	480
OY	481	catatctcgcctcggggagcctatttaataatacaacccctctagctgcctcatatgagtgctaaacc	540
Db	481	catatctcgcctcggggagcctatttaataatacaacccctctagctgcctcgaatgagtgctaaacc	540
OY	541	caaaaacccctactaaactgtgtgagtgtgctcccccctgtgaactctggagccatcatcttcaatc	600
Db	541	caaaaacccctactaaactgtgtgagtgtgctcccccctgtgaactctggagccatcatcttcaatc	600
OY	601	ccctttctcttgaaacaaatggaaaactcttaagcaacgaataaacaacactcttcggttttagta	660
Db	601	ccctttctcttgaaacaaatggaaaactcttaagcaacgaataaacaacactcttcggttttagta	660

[illegible]

Query Match	Best Local Similarity	100.0%	Score 1481	DB 20	Length 1481
Matches 1481	Conservative	0	Mismatches	0	Indels
			Gaps	0	
0Y	1	atggcctcccttaccatcattcttctcttactgtctcttacccttgcctcact	60		
Db	1	atggcctcccttaccatcattcttctcttactgtctcttacccttgcctcact	60		
0Y	61	gacccctccatagtgtgtgtacaaacagtagctcccttaccagaagttctatgaaga	120		
Db	61	gacccctccatagtgtgtgtacaaacagtagctcccttaccagaagttctatgaaga	120		
0Y	121	acggcgcttcctggagaatattgatgcccacatatagagtttctaaaggaaacc	180		
Db	121	acggcgcttcctggagaatattgatgcccacatatagagtttctaaaggaaacc	180		
0Y	181	accttactgcccacaccatattgcccgcgaactgtctaaactctgcactcttgcag	240		
Db	181	accttactgcccacaccatattgcccgcgaactgtctaaactctgcactcttgcag	240		
0Y	241	catgcaaatactcaattattgagcagaggaaatgatattcaactagtcttgcaggagact	300		
Db	241	catgcaaatactcaattattgagcagaggaaatgatattcaactagtcttgcaggagact	300		
0Y	301	ggagccactgtctgttgactacttaccaccatcacagtatgtctcgaaggaggatg	360		
Db	301	ggagccactgtctgttgactacttaccaccatcacagtatgtctcgaaggaggatg	360		
0Y	361	caaggctcagcagaagaagaacaagtaaaggagaacttcccaactgacccggggaact	420		
Db	361	caaggctcagcagaagaagaacaagtaaaggagaacttcccaactgacccggggaact	420		
0Y	421	agcacccctagccctacaaagagactagtctctctaaactacatgatgaaccctcg	480		
Db	421	agcacccctagccctacaaagagactagtctctctaaactacatgatgaaccctcg	480		
0Y	481	catattgcccgtgtagactatttataaccaccctcactggcttcacagagttccagc	540		
Db	481	catattgcccgtgtagactatttataaccaccctcactggcttcacagagttccagc	540		
0Y	541	caaaacccctactaaactgttggatgagctcccccctgcacttcaggcacaattccaat	600		
Db	541	caaaacccctactaaactgttggatgagctcccccctgcacttcaggcacaattccaat	600		
0Y	601	cctgttcctgaacaatgtaacaacttcagcacagaataaacaacacttcggttttaga	660		
Db	601	cctgttcctgaacaatgtaacaacttcagcacagaataaacaacacttcggttttaga	660		

QY	661	ggacctctgtttccaaatctcggaaataaccacaaactcaacctcaactgtgtataatt	720
Db	661	ggacctctgtttccaaatctcggaaataaccacaaactcaacctcaactgtgtataatt	720
QY	721	agcaatactaaagcaacaacgaactccaaatgtaagtgggttaaactctccacaga	780
Db	721	agcaatactaaagcaacaacgaactccaaatgtaagtgggttaaactctccacaga	780
QY	781	atagtctgcctaccctcaaggaataattttgtctgtgtgtacccatcatgtttg	840
Db	781	atagtctgcctaccctcaaggaataattttgtctgtgtgtacccatcatgtttg	840
QY	841	aatgctcttcagaactatctatgtgtcttcctcaatctcttagtgccccatagcaactac	900
Db	841	aatgctcttcagaactatctatgtgtcttcctcaatctcttagtgccccatagcaactac	900
QY	901	actaacaagaattatatacaatcatgttcgtatccaaagccccacaaagaagtaccatt	960
Db	901	actaacaagaattatatacaatcatgttcgtatccaaagccccacaaagaagtaccatt	960
QY	961	cttccttttgttatacgaagcagaggtgtctagcagactagctatgtgcattgtgcagttac	1020
Db	961	cttccttttgttatacgaagcagaggtgtctagcagactagctatgtgcattgtgcagttac	1020
QY	1021	acaacctctactcagttctctactacaactatctccaagaataatgtgtacaatgtgaacag	1080
Db	1021	acaacctctactcagttctctactacaactatctccaagaataatgtgtacaatgtgaacag	1080
QY	1081	gtcactgactccctcgtgtcacctcttgcaagaatcaacttaactctctagcagcagtagctctt	1140
Db	1081	gtcactgactccctcgtgtcacctcttgcaagaatcaacttaactctctagcagcagtagctctt	1140
QY	1141	caaaatcgaagaagcttttagactctgtctaaacccgcaaaagaagggggaacctgttatcttta	1200
Db	1141	caaaatcgaagaagcttttagactctgtctaaacccgcaaaagaagggggaacctgttatcttta	1200
QY	1201	ggagaagaagaagctgttattttgttaatcaatccagaattgtcaactgtgagaagtaaga	1260
Db	1201	ggagaagaagaagctgttattttgttaatcaatccagaattgtcaactgtgagaagtaaga	1260
QY	1261	attcgagatcgaaatacaaatgtagaagcagaagagcttcaaaaacgaacgcgtgggcttc	1320
Db	1261	attcgagatcgaaatacaaatgtagaagcagaagagcttcaaaaacgaacgcgtgggcttc	1320
QY	1321	ctcagccaatgtgatacccttgggtgtctccctctttagaagcctctagcagctctaaatgt	1380
Db	1321	ctcagccaatgtgatacccttgggtgtctccctctttagaagcctctagcagctctaaatgt	1380
QY	1381	ttactctcttttgagaccctgtatctttaacctctctgttaagtgtgtctctccagaatt	1440
Db	1381	ttactctcttttgagaccctgtatctttaacctctctgttaagtgtgtctctccagaatt	1440
QY	1441	gaagctgtaaagctacagatgctgtcttaacaatgtgaaccca	1481
Db	1441	gaagctgtaaagctacagatgctgtcttaacaatgtgaaccca	1481
RESULT 3			
ID AAA63826			
AAA63826 standard; DNA; 2030 BP.			
AC AAA63826;			
XX 04-DEC-2000 (first entry)			
XX Nucleotide sequence of the MSRV-1 3' env and LTR regions.			
DE MSRV-1; pol region; long terminal repeat; LTR; RUS region; retrovirus;			
XX ss.			
XX Multiple sclerosis retrovirus 1.			
XX			

```

FH Key Location/Qualifiers
FT CDS 1..1629
FT sig_peptide /note="Contains one termination codon"
FT 1..81
FT CAAAT_signal /tag="b"
FT 1800..1807
FT CAAAT_signal /tag="c"
FT 1858..1864
FT TATA_signal /tag="d"
FT 1906..1911
FT polyA_signal /tag="e"
FT 1996..2002
FT /tag="f"

XX WO20047745-A1.
XX 17-AUG-2000.
XX 15-FEB-2000; 2000WO-1B00159.
XX 15-FEB-1999; 99EP-0420041.
XX (INMR) BIO MERIEUX.
XX Paranhos-Baccala G, Perron H, Komurian-Pradel F;
XX WPI, 2000-506097/45.
XX DR P-PSDB: AAB08195.
XX Nucleotide fragment of LTR-RUS region from Multiple Sclerosis
XX retrovirus (MSRV) used to detect the presence of MSRV-1 retrovirus in a
XX biological sample
XX
XX Disclosure: Fig 2; 23pp; English.
XX
XX The present sequence represents the nucleotide sequence corresponding
XX to the 3' env region and long terminal repeat sequences from clone
XX CL6 of Multiple Sclerosis retrovirus (MSRV-1). The specification
XX describes a long terminal repeat (LTR)-RUS region which encodes the
XX expression of a MSRV-1 protein. This is unusual for LTRs, in
XX particular in the RUS region. The sequence includes CAAAT and TATA
XX signals which are present in the U3 and R regions and are not directed
XX towards the CDS indicated in the features table. Probes and antibodies
XX to the MSRV-1 retrovirus protein and encoding polynucleotide sequences
XX are used to detect the presence of MSRV-1 retrovirus in a biological
XX sample.
XX
XX Sequence 2030 BP; 574 A; 559 C; 387 G; 510 T; 0 other:
XX
XX Query Match 100.0%; Score 1481; DB 21; Length 2030;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 1481; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 241 catgcaataactcaattattgagcagggaataatgataatccagtgtgtcctgagagactt 300
Qy 301 ggagccactgtctgtgagacttacttaccacataacagatgtctgagtgggtggaatt 360
Db 301 ggagccactgtctgtgagacttacttaccacataacagatgtctgagtgggtggaatt 360
Qy 361 caaggtcagcgaagagagaaacaagtaagaagcaatctccaaactgacccgggagat 420
Db 361 caaggtcagcgaagagagaaacaagtaagaagcaatctccaaactgacccgggagat 420
Qy 421 agcaccctagcccttaccagaagagactagttctcctcaaaactcaatgaaacccctgtacc 480
Db 421 agcaccctagcccttaccagaagagactagttctcctcaaaactcaatgaaacccctgtacc 480
Qy 481 catactcgtctgtgagccatttaataacacccctcactcgtcctcatgaggtctcagcc 540
Db 481 catactcgtctgtgagccatttaataacacccctcactcgtcctcatgaggtctcagcc 540
Qy 541 caaaacccactaactgttggatgtgctccctcgtcacttcaaggcatalacttcaatc 600
Db 541 caaaacccactaactgttggatgtgctccctcgtcacttcaaggcatalacttcaatc 600
Qy 601 cctgttctcgaacaatgagacacacttcaagacagaataaacaacacttcggttttagta 660
Db 601 cctgttctcgaacaatgagacacacttcaagacagaataaacaacacttcggttttagta 660
Qy 661 ggacctctgttcccaatcctggaataaccataccataccactcactcgtgttaaat 720
Db 661 ggacctctgttcccaatcctggaataaccataccataccactcactcgtgttaaat 720
Qy 721 agcaataactatgacacaacacagctcccaatgcatcaggttgggttaacacctccacaga 780
Db 721 agcaataactatgacacaacacagctcccaatgcatcaggttgggttaacacctccacaga 780
Qy 781 ataagctcgtcctccctcagaagaatttttgcgtgagacccctcagccatattgtttg 840
Db 781 ataagctcgtcctccctcagaagaatttttgcgtgagacccctcagccatattgtttg 840
Qy 841 aatgctctcctcagaactatgtgctcctcctcatcttcttgaatggccctcctacacatcac 900
Db 841 aatgctctcctcagaactatgtgctcctcctcatcttcttgaatggccctcctacacatcac 900
Qy 901 actgacaagaattatacaaatcatatgtctgaactaaagcccaacaagaagtaaccatt 960
Db 901 actgacaagaattatacaaatcatatgtctgaactaaagcccaacaagaagtaaccatt 960
Qy 961 cctccttctgtatatacagaacagagatgctgaagcaagctagtgactggtcaggtatc 1020
Db 961 cctccttctgtatatacagaacagagatgctgaagcaagctagtgactggtcaggtatc 1020
Qy 1021 acaacctactcagttctactactaacaactatctcaagaataaatggtgacatggagacag 1080
Db 1021 acaacctactcagttctactactaacaactatctcaagaataaatggtgacatggagacag 1080
Qy 1081 gtcactgactcctcgtgtcactcttgcagaatcaacttaactcctcagcagtagtccct 1140
Db 1081 gtcactgactcctcgtgtcactcttgcagaatcaacttaactcctcagcagtagtccct 1140
Qy 1141 caaaatcgaagaagcttgaactgtcgaacggccaaagaaggggggaacccgtttatttta 1200
Db 1141 caaaatcgaagaagcttgaactgtcgaacggccaaagaaggggggaacccgtttatttta 1200
Qy 1201 ggaagaagaagctgttatatgttlaatcaatccagaatgtgacatgagagaagttaagaaga 1260
Db 1201 ggaagaagaagctgttatatgttlaatcaatccagaatgtgacatgagagaagttaagaaga 1260
Qy 1261 attcgagatcgataacaatgtgtagagcagagagcttcaaaacacgacgtggggccctc 1320
Db 1261 attcgagatcgataacaatgtgtagagcagagagcttcaaaacacgacgtggggccctc 1320
Qy 1321 ctcagcgaatgtagtccctgggttctccctccttagagcccttagcagcttaatatg 1380
Db 1321 ctcagcgaatgtagtccctgggttctccctccttagagcccttagcagcttaatatg 1380

```

QY 1381 ttactcctttgaccctgacatcttaacctctgtttaagttgtctcttcagaatt 1440
|||||
Db 1381 ttactcctttgaccctgacatcttaacctctgtttaagttgtctcttcagaatt 1440
QY 1441 gaagctgtaagctacagatggtcttacaatggaaccca 1481
|||||
Db 1441 gaagctgtaagctacagatggtcttacaatggaaccca 1481
RESULT 4
AAA96625 standard; DNA: 1629 BP.
XX
AC AAA96625;
XX
DT 08-FEB-2001 (first entry)
XX
DE DNA encoding an envelope (env) protein of MSRV-1.
XX
NM Envelope protein: MSRV-1; superantigen; autoimmune disease; Vbeta16;
NM Vbeta17; multiple sclerosis; vaccine; ss.
XX
OS Human endogenous retrovirus.
XX
FH Key Location/Qualifiers
FT CDS 1..1629
FT /*tag= a
FT /product= "envelope protein"
XX
PN MO200057185-A1.
XX
PD 28-SEP-2000.
XX
PF 20-MAR-2000; 2000MO-FR00691.
XX
PR 19-MAR-1999; 99FR-0003622.
PR 28-OCT-1999; 99FR-0013755.
XX
PA (INMR) BIO MERIEUX.
XX
PI Perron H, Lafont M;
XX
DR WPI: 2000-638279/61.
DR P-PSDB; AAB19069.
XX
PT Detecting superantigen activity, useful for identifying agents for
PT treatment or prevention of autoimmune disease, from expansion or loss
PT of particular lymphocyte Vbeta determinants -
XX
PS Claim 25; Page 123-124; 134pp; French.
XX
XX The present sequence encodes an envelope protein of MSRV-1. The
XX envelope protein expressed by the endogenous human retrovirus MSRV-1
XX has superantigen activity associated with autoimmune disease. The
XX protein can be detected using the method of the invention. The
XX specification describes a process for detecting activity of a
XX superantigen in a biological sample. The process comprises identifying
XX large scale expansion or loss of lymphocytes that carry at least one
XX of the Vbeta16 and/or Vbeta17 determinants. The method is used to
XX screen for agents that inhibit the superantigen, especially those
XX associated with MSRV-1 which is implicated in autoimmune disease,
XX particularly multiple sclerosis. These agents are potentially useful
XX for treatment or prevention (e.g. as vaccines) of autoimmune diseases.
XX
SQ Sequence 1629 BP; 456 A; 454 C; 287 G; 432 T; 0 other;

Query Match 98.7%; Score 1461.8; DB 21; Length 1629;
Best Local Similarity 99.2%; Pred. No 0;
Matches 1469; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 atggccctccctatcatattcttcttactgtctcttacccttcgtctcact 60

Db 1 atggccctccctatcatattcttcttactgtctcttacccttcgtctcact 60
|||||
QY 61 gcaacccctcctatctgtctgtacaaacagtagtgcctcccttccaaagattctatgaaga 120
|||||
Db 61 gcaacccctcctatctgtctgtacaaacagtagtgcctcccttccaaagattctatgaaga 120
QY 121 acgagctctcttgaaatgatgagcccatcatataggaattatctaaaggaaaccc 180
|||||
Db 121 acgagctctcttgaaatgatgagcccatcatataggaattatctaaaggaaaccc 180
QY 181 accttcaatgccacacacatattgcccgcgaactgtataactctgcaactcttgcattg 240
|||||
Db 181 accttcaatgccacacacatattgcccgcgaactgtataactctgcaactcttgcattg 240
QY 241 catgcaaatcatcatattatgacagggaaatgatactcctagtgtctctgagagact 300
|||||
Db 241 catgcaaatcatcatattatgacagggaaatgatactcctagtgtctctgagagact 300
QY 301 ggaagccatgctgttctgacttacttaccatccagtagtctgtatgggggtggaatt 360
|||||
Db 301 ggaagccatgctgttctgacttacttaccatccagtagtctgtatgggggtggaatt 360
QY 361 caaggtcagcagaagaaacaagtaagaagaatctcccaactgtaccgggagacat 420
|||||
Db 361 caaggtcagcagaagaaacaagtaagaagaatctcccaactgtaccgggagacat 420
QY 421 agcacccctagcccttacaagaagactgtctctcoaaactacatgaaacccctcgtacc 480
|||||
Db 421 agcacccctagcccttacaagaagactgtctctcoaaactacatgaaacccctcgtacc 480
QY 481 catatcgctggtgtagcttattatatacccttccctactggtccatagaagttctcagcc 540
|||||
Db 481 catatcgctggtgtagcttattatatacccttccctactggtccatagaagttctcagcc 540
QY 541 caaaacccctactaactgttgaatgtgctccctccctgcaactcagggccatacttccaatc 600
|||||
Db 541 caaaacccctactaactgttgaatgtgctccctccctgcaactcagggccatacttccaatc 600
QY 601 cctgttctgaaacaatggaacaacttcagcaagaagaataaacaacactcctggtttagta 660
|||||
Db 601 cctgttctgaaacaatggaacaacttcagcaagaagaataaacaacactcctggtttagta 660
QY 661 ggaactctgttcccaatctggaataaccatccctcaaacctccactcgtgtataatt 720
|||||
Db 661 ggaactctgttcccaatctggaataaccatccctcaaacctccactcgtgtataatt 720
QY 721 agcaatactatagacacacacagctcccaatgcatcaggtgtgtaacaccccccacaga 780
|||||
Db 721 agcaatactatagacacacacagctcccaatgcatcaggtgtgtaacaccccccacaga 780
QY 781 atagtctgcttaacctcagaatatatttctgtgtgtgtactcgaagctatcatgtttg 840
|||||
Db 781 atagtctgcttaacctcagaatatatttctgtgtgtgtactcgaagctatcatgtttg 840
QY 841 aatggcttctaagaatcatgtgtctctcctcatattcttaagtcctccatatacctac 900
|||||
Db 841 aatggcttctaagaatcatgtgtctctcctcatattcttaagtcctccatatacctac 900
QY 901 actgaacaagattatatacaatcatgtcttaactaaagccccaacaaagaagtagtaccat 960
|||||
Db 901 actgaacaagattatatacaatcatgtcttaactaaagccccaacaaagaagtagtaccat 960
QY 961 ctctccttgttatcagagcagagtgctagggcagaactagtagtgcattggcagatc 1020
|||||
Db 961 ctctccttgttatcagagcagagtgctagggcagaactagtagtgcattggcagatc 1020
QY 1021 acaaccttactcagttcttactatacaactatctcaagaataaatagtgtcagtgaaag 1080
|||||
Db 1021 acaaccttactcagttcttactatacaactatctcaagaataaatagtgtcagtgaaag 1080
QY 1081 gtcaatgactccctgtgtaaccttgcaagatacaacttaactcctcctagcagcagtagtctt 1140
|||||

[illegible]

XX		HERV-W envelope protein G coding sequence.
DE		
XX		
KM		Human endogenous retrovirus; HERV-W; HERV; chromosome 7; env protein;
KW		envelope protein; multiple sclerosis-related superantigen; vaccine;
KM		surface antigen; transmembrane; multiple sclerosis; neuroprotective;
KW		antisense-therapy; autoimmune disorder; ds.
XX		
OS		Human endogenous retrovirus.
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..1617
FT	/tag=	a
FT	/product=	"HERV-W envelope protein G"
PN		WO200131021-A1.
XX		
PD		03-MAY-2001.
XX		
PF		30-OCT-2000; 2000MO-EP10659.
XX		
PR		28-OCT-1999; 99EP-0402690.
XX		
PA		(UYGE-) UNIV GENEVE.
PI		Conrad B, Mach B;
XX		
DR		WPI: 2001-316336/33.
XX		P-PsDB; AAB75138.
PT		New human retrovirus HERV-W ENV proteins/peptides having superantigen
PS		activity useful for diagnosing and treating multiple sclerosis -
XX		Claim 13; Fig 10; 94pp; English.
CC		
CC		On the basis of the PBS t-RNA motif used for the classification of human
CC		endogenous retrovirus (HERVs) the full length endogenous provirus which
CC		was been located on the long arm of human chromosome 7 (7q21-22) has been
CC		designated HERV-W. The present invention describes proteins or peptides
CC		(I) having superantigen (SAG) activity comprising the ENV protein (ENV)
CC		of HERV-W, the surface protein (SU) and transmembrane (TM) sub-units. (I)
CC		have neuroprotective activity, and can be used in: vaccines; antisense-
CC		therapy; and HERV-W SAG activity-inhibitors. (I) and encoding DNA/RNA are
CC		useful for diagnosing multiple sclerosis (MS) or HERV-W-associated
CC		disorders. (I) are also useful for identifying substances (and optionally
CC		recovering) capable of binding to a retroviral superantigen associated
CC		with MS, substances capable of blocking SAG activity and substances
CC		capable of blocking transcription or translation of HPRV-W retroviral
CC		superantigen. A protein or peptide derived from (I), modified to be
CC		devoid of SAG activity and being capable of generating an immune response
CC		against HERV-W retroviral SAG is useful in therapy. Nucleic acid
CC		molecules encoding (I) are useful as vaccines against MS. Substances
CC		capable of blocking SAG activity, capable of binding to a retroviral
CC		superantigen associated with MS, or capable of blocking transcription or
CC		translation of HERV-W retroviral superantigen for use in treating or
CC		preventing MS, obtained using (I) are useful for the treatment and
CC		prevention of MS. (I) and nucleic acids encoding them are useful for
CC		diagnosing autoimmune disease. The present sequence encodes the
CC		specifically claimed envelope protein of HERV-W designated G.
XX		
SQ		Sequence 1617 BP; 442 A; 452 C; 296 G; 427 T; 0 other:
Query Match	90.1%;	Score 1333.8; DB 22; Length 1617;
Best Local Similarity	93.8%;	Pred. No. 0;
Matches 1389; Conservative	0;	Mismatches 92; Indels 0; Gaps 0;
OY		1 atggcccccttaccatcattcttccttactggtttccctacccccttgctcact 60
db		1 atggcccccttaccatcattcttccttactggtttccctacccccttcacctcact 60
		gcacccccctcacatgctgctgtacaacagctagctcccttaccagaagtctcatgaaga 120

Db 61 gacccccctcctacgctgtatgacagtagtccctaccagaagtctctatggaga 120
 QY 121 acgagctctccctgaaatatgtatgcccccatatagagagttatctaaaggaaatcc 180
 Db 121 atgacagctcccgaaatatatgtatgcccccatgtatagaggtctctttaaaggaaatcc 180
 QY 181 acctactgcccacacacccatgccccgaactgtatatacttctgcaactcttggatg 240
 Db 181 acctactgcccacacacccatgccccgaactgtatatacttctgcaactcttggatg 240
 QY 241 catgcaaatatcatatatttgacagaggaataatattatccctagtctccttgagagact 300
 Db 241 catgcaaatatcatatatttgacagaggaataatattatccctagtctccttgagagact 300
 QY 301 gggagccactgctgttgacttacttaccacataccagtagtctgtatgggggtgagatt 360
 Db 301 gggagccactgctgttgacttacttaccacataccagtagtctgtatgggggtgagatt 360
 QY 361 caagctcagcgaaggaanaaacaagtaaggaacaaatctcccaactgaccgggagacat 420
 Db 361 caagctcagcgaaggaanaaacaagtaaggaacaaatctcccaactgaccgggagacat 420
 QY 421 agacccccctgccccctacaaagactagctctcaaaactacatgaacccctcgatcc 480
 Db 421 agacccccctgccccctacaaagactagctctcaaaactacatgaacccctcgatcc 480
 QY 481 catactcgcctgtgtagcctatttaatacacccctcactgctgacatgaggtctagacc 540
 Db 481 catactcgcctgtgtagcctatttaatacacccctcactgctgacatgaggtctagacc 540
 QY 541 caaaacccctactactgctgtgatgtgctcccccctgacactcagggccatcatcattc 600
 Db 541 caaaacccctactactgctgtgatgtgctcccccctgacactcagggccatcatcattc 600
 QY 601 cctgttctcgaacaatggaacaactctcagcagaataaacaacactccgttttagta 660
 Db 601 cctgttctcgaacaatggaacaactctcagcagaataaacaacactccgttttagta 660
 QY 661 ggaacctctgttcccaatctggaataacccatcccaactcactcactgtgttaaat 720
 Db 661 ggaacctctgttcccaatctggaataacccatcccaactcactcactgtgttaaat 720
 QY 721 agcaatactatagacaacaacagctcccaatgcatcagtggtgaacaactccacacga 780
 Db 721 agcaatactatagacaacaacagctcccaatgcatcagtggtgaacaactccacacga 780
 QY 781 atagctcgtcctacccctaggaataattttgtctgtgtacccctacatcatgtttg 840
 Db 781 atagctcgtcctacccctaggaataattttgtctgtgtacccctacatcatgtttg 840
 QY 841 aatgagctctcagaatctatgtctcctcctcacttactttagtccccctatgacatctac 900
 Db 841 aatgagctctcagaatctatgtctcctcctcacttactttagtccccctatgacatctac 900
 QY 901 actgaaacaagattatatacaatctgctacctaagccccacaacaaagatgacat 960
 Db 901 actgaaacaagattatatacaatctgctacctaagccccacaacaaagatgacat 960
 QY 961 ctctcctttgttatcagagcagagtgctagggcagactaggtactgcatgtgacatc 1020
 Db 961 ctctcctttgttatcagagcagagtgctagggcagactaggtactgcatgtgacatc 1020
 QY 1021 acaacctctactcagttctctactaacaactctcacaagaataatgtgtgacatgagacag 1080
 Db 1021 acaacctctactcagttctctactaacaactctcacaagaataatgtgtgacatgagacag 1080
 QY 1081 gtcactgactccctgctgacactgtgcaagatcaacttactcctcctagagcagtagtctc 1140
 Db 1081 gtcgagcagccctgctgacactgtgcaagatcaacttactcctcctagagcagtagtctc 1140
 QY 1141 caaatcgaagagcttagactgtctacacgccaagaagggggaacctgttatcttta 1200
 Db 1141 caaatcgaagagcttagactgtctacacgccaagaagggggaacctgttatcttta 1200

QY 1201 ggaagaagacgtgttatgtttaatcaatccagaattgtcactggaagaaatgaaga 1260
 Db 1201 ggggaagaagacgtgttatgtttaatcaatccagaattgtcactggaagaaatgaaga 1260
 QY 1261 atcagagatcgaatatacaatgtlagagcagagagcttcaaacacccagacgtgggctc 1320
 Db 1261 atcagagatcgaatatacaatgtlagagcagagagcttcaaacacccagacgtgggctc 1320
 QY 1321 ctgagccaatgtagtgcctgtgctccctctttagagacctcagcagctataattg 1380
 Db 1321 ctgagccaatgtagtgcctgtgctccctctttagagacctcagcagctataattg 1380
 QY 1381 ttactcctcttgagcccttatcttatacctcctgttgaattgtctctccagaat 1440
 Db 1381 ctactcctcttgagcccttatcttatacctcctgttgaattgtctctccagaat 1440
 QY 1441 gaagctgtaaagctacagatggtcttacaatgtgaacccca 1481
 Db 1441 gaagctgtaaactacaatgtgagcccaagatgagtcacaa 1481

RESULT 8
 AAF55630
 ID AAF55630 standard; DNA; 2781 BP.
 XX
 AC AAF55630;
 DT 29-MAY-2001 (first entry)
 XX
 DE Nucleotide sequence of a human endogenous retrovirus envelope protein.
 XX
 KW Envelope protein; HERV; syncytia formation; placental development;
 KM syncytia; cancer; cell adhesion; ss.
 OS Human endogenous retrovirus.
 XX
 FH Key Location/Qualifiers
 FT CDS 762..2378
 FT /tag= a
 FT /product= "envelope protein"
 XX
 XX WO200116171-A1.
 PD 08-MAR-2001.
 XX
 PF 01-SEP-2000; 2000MO-FR02429.
 XX
 PR 01-SEP-1999; 99FR-0011141.
 PR 15-SEP-1999; 99FR-0011793.
 XX
 PA (TNMR) BIO MERIEUX.
 PA (TNMR) INST NAT SANTE & RECH MEDICALE.
 XX
 PI Maillet F, Cosset F, Blond J, Lavillette D, Bouton O, Ruggieri A;
 DR WPI: 2001-226676/23.
 DR P-PSDB; AAB67652.
 XX
 PT Detecting expression of human endogenous retrovirus envelope protein in
 PS cells of a tissue or culture, from its ability to induce syncytia -
 XX
 PS Disclosure: Page 44-45; 57pp; French.
 XX
 CC The present sequence encodes a human endogenous retrovirus envelope
 CC protein. The specification describes a method for detecting expression
 CC of an envelope protein from a human endogenous retrovirus (HERV), in
 CC cells, of a tissue or culture. The method comprises detecting syncytia
 CC formation due to the fusogenic properties of the envelope protein.
 CC Envelope polypeptides and polynucleotides are used to produce
 CC therapeutic or prophylactic compositions, particularly for treatment of
 CC cancer, to correct defects in placental development (or other natural
 CC formation of other types of syncytia), and to promote adhesion of cells

CC	of HERV-W, the surface protein (SU) and transmembrane (TM) sub-units. (1)
CC	have neuroprotective activity, and can be used in: vaccines; antisense-
CC	therapy; and HERV-W SAg activity-inhibitors. (1) and encoding DNA/RNA are
CC	useful for diagnosing multiple sclerosis (MS) or HERV-W-associated
CC	disorders. (1) are also useful for identifying substances (and optionally
CC	recovering) capable of binding to a retroviral superantigen associated
CC	with MS, substances capable of blocking SAg activity and substances
CC	capable of blocking transcription or translation of HERV-W retroviral
CC	superantigen. A protein or peptide derived from (1), modified to be
CC	devoid of SAg activity and being capable of generating an immune response
CC	against HERV-W retroviral SAg is useful in therapy. Nucleic acid
CC	molecules encoding (1) are useful as vaccines against MS. Substances
CC	capable of blocking SAg activity, capable of binding to a retroviral
CC	superantigen associated with MS, or capable of blocking transcription or
CC	translation of HERV-W retroviral superantigen for use in treating or
CC	preventing MS, obtained using (1) are useful for the treatment and
CC	prevention of MS. (1) and nucleic acids encoding them are useful for
CC	diagnosing autoimmune disease. The present sequence encodes the
CC	specifically claimed envelope protein of HERV-W designated G.
SQ	Sequence 2782 BP; 741 A; 768 C; 564 G; 709 T; 0 other;
Query Match	88.3%; Score 1308.2; DB 22; Length 2782;
Best Local Similarity	92.7%; Pred. No. 0;
Matches 1373; Conservative	0; Mismatches 108; Indels 0; Gaps 0;
QY	1 atggccctccctatcatcatttcttccttaactgttctctaaccccccttgctctaact 60
DB	763 atggccctccctatcatcatttcttccttaactgttctctaaccccccttgctctaact 822
OY	61 gaccccctccatgctgtcgtaacaacagtagctccctccaagaagtctcatagaaga 120
DB	823 gaccccctccatgctgcgtgtatgatgacagctgctccctaccaggattcctatgaga 882
OY	121 acgcygcttcctgtaaatattgatgcccatcataatagagtttatctaaaggaaaactcc 180
DB	883 atgcagcgtcccggaatatattgatgcccatcgatagtagagttctttaaggaaacccc 942
OY	181 acctcaactgcccacaacccataatgccccggaactgctataactctgccaactcttgatg 240
DB	943 accttcaactgcccacaacccataatgccccggaactgctataactctgccaactcttgatg 1002
OY	241 catgtaaatattcatattatgtagcaggaggaatagattaacctagttgctccggagaact 300
DB	1003 catgtaaatattcatattatgtagcaggaggaatagattaacctagttgctccggagaact 1062
OY	301 ggaagcacatgtctgttgactactctcaaccaatcaaaglatgctcgaatgggggtgaatt 360
DB	1063 ggaagcacatgtctgttgactactctcaaccaactgtagtgtctgatalgggggtgagtt 1122
OY	361 caaagtcaggcagaagaaaaacaagtaagaagaagcaatctcccaactgacccgggagat 420
DB	1123 caaagtcaggcagaagaaaaacatlttaaagaagaatctcccaactcgaacgggtatcat 1182
OY	421 agcacccctaggcccttaaaagagactagtctcttaaaactatacagaaacctccgtacc 480
DB	1183 agcacccctaggcccttaaaagagactagatctcttaaaactatacagaaacctccgtacc 1242
OY	481 cattaactgcctgtgtagcctatttaataccaacccctcaactcggctcagaaggtccagcc 540
DB	1243 cattaactgcctgtgtagcctatttaataccaacccctcaactcggctcagaaggtccagcc 1302
OY	541 caaaacccataactagtgtgagtgtgctccctcgcaactlccagggcatalacattcaatc 600
DB	1303 caaaacccataactagtgtgagtgtgctccctcgcaactlccagggcatalacattcaatc 1362
OY	601 cctgttcctgaaacaatgnaacaacttgagacaagaataaacaacaacttcogttttaata 660
DB	1363 cctgttacccggaacaatgnaacaacttcagcacaagaataaacaacaacttcogttttaata 1422
OY	661 gaacctctgtttccaatctggaataaccccaacctcaaacactcactgtgttaaatct 720

Db	1423	ggaacctctgtttccaaatgtggaataataaccataaccataaccataaccataaccctgtgttaaatctt	1482
Qy	721	agcaaatctatagaagacaacaacagcctccccaatgcataggtgtgtgtatacaacctccacaaga	780
Db	1483	agcaaatctatcatacaacaacaacacctcccaatgcataggtgtgtgtatacaacctccacaaca	1542
Qy	781	ataatctgcgcctaacccctcaggaataatctttgtctgtgtgtgtgtgtgtgtgtgtgtgtgt	840
Db	1543	ataatctgcgcctaacccctcaggaataatctttgtctgtgtgtgtgtgtgtgtgtgtgtgtgt	1602
Qy	841	aaatgctctctcagaatctaatgtgtctctctctcatcttcttaagtgtccccctatgaccatctac	900
Db	1603	aaatgctctctcagaatctaatgtgtctctctctcatcttcttaagtgtccccctatgaccatctac	1662
Qy	901	actgtaaacagaatcttatctaatctatctgtctgtacttaagcccccaaaaaaagaatgccact	960
Db	1663	actgtaaacagaatcttatctaatctatctgtctgtacttaagcccccaaaaaaagaatgccact	1722
Qy	961	ctctccctttgtatcaagaagagagatgtcttaagagaagctagtgactgtgtgtgtgtgtgtgt	1020
Db	1723	ctctccctttgtatcaagaagagagatgtcttaagagaagctagtgactgtgtgtgtgtgtgtgt	1782
Qy	1021	acaaacctctacccagatgtcttactctacaacatactccaagaataaaltgtgtacatgtgaacag	1080
Db	1783	acaaacctctacccagatgtcttactctacaacatactccaagaataaaltgtgtacatgtgaacag	1842
Qy	1081	gtcacgtgacccctccctgtgtcacactctgtcagaagatctaaactaaactccctctgacagcagtagctct	1140
Db	1843	gtcacgtgacccctccctgtgtcacactctgtcagaagatctaaactaaactccctctgacagcagtagctct	1902
Qy	1141	caaaatctgaagaagatctttaagctctgtcccaacgcgcgaagaagggggagacccgtttatttcta	1200
Db	1903	cgaaatctgaagaagatctttaagctctgtcccaacgcgcgcgaagaagggggagacccgtttatttcta	1962
Qy	1201	qgagaaagacgctgtttatttgttttaactcaatcccgaaatgttcaactgtgaaagtttaagaa	1260
Db	1963	ggggaagaatgtctgtttatttgttttaactcaatcccgaaatgttcaactgtgaaagtttaagaa	2022
Qy	1261	atctcgaaatctgaaatacaatgtttaagagaagagggtcttcaaaaacacgcgaacgcctggggcctc	1320
Db	2023	atctcgaaatctgaaatacaacgtaataagagaagagctctcgaaaaacacgtggacccctggggcctc	2082
Qy	1321	ctcgcgcgaatggaatgcctctggctgtctcccccctcttgaagaagcctctgagccgtcttaatatgt	1380
Db	2083	ctcgcgcgaatggaatgagccctgtgattctcccccctcttgaagaagcctctgagccgtcttaatatgt	2142
Qy	1381	ttactctctcttggaaacctgtatctttaaacctctctgttaagtgtgtctctccagaagt	1440
Db	2143	ctactctctcttggaaacctgtatctttaaacctctctgttaagtgtgtctctccagaatc	2202
Qy	1441	gaagctgttaagcttaacgaatgagctcttaacaaatgtgaaccccca	1481
Db	2203	gaagctgttaagcttaacaaatgtgaagcccaagaatgtcagttccaa	2243

PD 21-JAN-1999.
XX
XX 06-JUL-1998; 98WO-FR01442.
PF XX
XX 07-JUL-1997; 97FR-0008815.
PR XX
XX (INMR) BIO MERIEUX.
PA
PI Beseme F, Blond JL, Boulton O, Mallet F, Mandrand B;
XX
XX WPI; 1999-120897/10.
DR
XX New nucleic acid sequences from human endogenous retrovirus-W -
PT expressed exclusively in placenta and useful in diagnosis and
PT therapy of autoimmune disease, and abnormal or failed pregnancy
XX
PS Claim 1; Page 60-63; 106pp; French.

This sequence represents clone cI.PH74 of the human endogenous retrovirus (HERV) W genome. The nucleic acids, their fragments or peptides encoded by them are markers of autoimmune disease (e.g. multiple sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-dependent diabetes and related pathologies) and of abnormal or unsuccessful pregnancy and can be used as chromosomal markers for susceptibility to these conditions, or proximity markers of genes associated with this susceptibility.

Sequence 2782 BP; 741 A; 767 C; 565 G; 709 T; 0 other:

Query Match	88.2%	Score 1306.6	DB 20	Length 2782
Best Local Similarity	92.6%	Pred. No. 0		
Matches 1372	Conservative	0	Mismatches 109	Indels 0
			Gaps	0

[illegible]

QY 601 cctgttcctgaacaatggaacaacttcagcacagaataaaccactctccgttttagta 660
 DB 1363 cctgttactgaacaatggaacaacttcagcacagaataaaccactctccgttttagta 1422
 QY 661 ggaacctctgttcccaatctggaataaaccatacctcaaacctcactctgtgtataatt 720
 DB 1423 ggaacctctgttcccaatctggaataaaccatacctcaaacctcactctgtgtataatt 1482
 QY 721 agcaatactatagaacaacaagctcccaatgcatcaggggtggaataacacctccacacga 780
 DB 1483 agcaatactatagaacaacaagctcccaatgcatcaggggtggaataacacctccacacga 1542
 QY 781 ataagtcgctacacccaagaataattttgtctgtgtgtacctagacctatcatgttttg 840
 DB 1543 ataagtcgctacacccaagaataattttgtctgtgtgtacctagacctatcatgttttg 1602
 QY 841 aatgtgctcttcagaataatctgtgtctccctcattctttagtgcctccatgacacatcac 900
 DB 1603 aatgtgctcttcagaataatctgtgtctccctcattctttagtgcctccatgacacatcac 1662
 QY 901 actgaaacaagattatatacaatcgtgtacttaagcccccaacaagaagatccatt 960
 DB 1663 actgaaacaagattatatacaatcgtgtacttaagcccccaacaagaagatccatt 1722
 QY 961 ctctcctttttatcagagcagaggtgtcagcagactaaggtactgtgcatgtgcagatc 1020
 DB 1723 ctctcctttttatcagagcagaggtgtcagcagactaaggtactgtgcatgtgcagatc 1782
 QY 1021 acaacctctactcagttctactacaactatcacaagaataaattgtgtacatgtgaacag 1080
 DB 1783 acaacctctactcagttctactacaactatcacaagaataaattgtgtacatgtgaacag 1842
 QY 1081 gtctctgactcctcgtggtcactctgcaagatcaacttaactcctcagcagcagatgtcctt 1140
 DB 1843 gtctcgcagactcctcgtggtcactctgcaagatcaacttaactcctcagcagcagatgtcctt 1902
 QY 1141 caaatacgaagaagctttagactgtctaaccgcaagaagagggggaacctgtttatttta 1200
 DB 1903 cgaatacgaagaagctttagactgtctaaccgcaagaagagggggaacctgtttatttta 1962
 QY 1201 ggagaagaagcgtgttattatgttaatcaatccagaattgtcactgtgagaagaattaaaga 1260
 DB 1963 ggagaagaagcgtgttattatgttaatcaatccagaattgtcactgtgagaagaattaaaga 2022
 QY 1261 attgagaatcgaataatgagagcagagagcttcaaaacaccggaacgtgtgggctc 1320
 DB 2023 attcagaatcgaataatgagagcagagagcttcaaaacaccggaacgtgtgggctc 2082
 QY 1321 ctcaagcaatgagatgacctgtgtctccctctcttagagacctcagacgctctaataatg 1380
 DB 2083 ctcaagcaatgagatgacctgtgtctccctctcttagagacctcagacgctctaataatg 2142
 QY 1381 ttaactcctcttggaaacctgtatcttaacctcctgtttaaagttgtctctccagaatt 1440
 DB 2143 ctactcctcttggaaacctgtatcttgaacctcctgtttaaagttgtctctccagaatt 2202
 QY 1441 gaaagctgtaaagcagagatgtcttacaagtgaaccca 1481
 DB 2203 gaaagctgtaaagcagagatgtcttacaagtgaaccca 2243

RESULT 13

AAA59211
 ID AAA59211 standard; DNA; 2782 BP.

AC AAA59211;

DT 07-NOV-2000 (first entry)

DE 5' non coding, 3' pol gene, and 3' non coding sequences of HERV-W.

XX Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;

KW gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
 XX Homo sapiens.
 OS
 XX
 XX MO200043521-A2.
 PN
 XX
 PD 27-JUL-2000.
 XX
 PF 21-JAN-2000; 2000MO-FR00144.
 XX
 PR 21-JAN-1999; 99PR-0000888.
 XX
 PA (INMR) BIO MERIEUX.
 PI Paranhos-Baccala G, Mallet F, Voisset C;
 XX
 DR WPI: 2000-499229/44.
 XX
 PT New nucleic acid from human endogenous retrovirus, useful e.g. for
 PT diagnosis of autoimmune disease and complications of pregnancy,
 PT contains at least part of the gag gene -
 PS Disclosure; Page 46-47; 53pp; French.
 XX
 CC The present sequence represents an endogenous retroviral nucleic acid
 CC fragment, which is associated with an autoimmune disease, and is
 CC integrated into the human genome. The fragment is originally derived
 CC from a novel retrovirus, human endogenous retrovirus W (HERV-W). The
 CC HERV-W retrovirus is associated with autoimmune disease, failure of
 CC pregnancy or disorders of pregnancy. The nucleic acid fragment, or
 CC proteins derived from it, are useful for diagnosis of autoimmune
 CC disease (specifically multiple sclerosis) and for monitoring pregnancy.
 CC The nucleic acid fragments may also be used for in situ labelling of
 CC isolated chromosomes, while the transcription product can be used to
 CC study or monitor T cell proliferation in vitro.
 XX
 SQ Sequence 2782 BP; 741 A; 767 C; 565 G; 709 T; 0 other;

Query Match 88.2%; Score 1306.6; DB 21; Length 2782;

Best Local Similarity 92.6%; Pred. No. 0;

Matches 1372; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 1 atggccctcccttatactctctcttactgttctcttaccctccctcctcact 60
 DB 763 atgggctcccttatactctctcttactgttctcttaccctcctcact 822
 QY 61 gcaacccctcattgtctgtacacaacagtagctcccttaccacaagattctatgaaga 120
 DB 823 gcaacccctcattgtctgtacacaacagtagctcccttaccacaagattctctatgaga 882
 QY 121 acgggctctctgtaaatatgatgccccatcatataggaattatctaaaggaaacccc 180
 DB 883 atgagcgttcccggaataatgatgccccatcatataggaattctctaaaggaaacccc 942
 QY 181 accttcatgcccacaccatatacgcccgcaactgctataactctgccaactcttgatga 240
 DB 943 accttcatgcccacaccatatacgcccgcaactgctataactctgccaactcttgatga 1002
 QY 241 catgcaatcatcattatgtgacaggggaatgttaactctagtgtgtcctggaagact 300
 DB 1003 catgcaatcatcattatgtgacaggggaatgttaactctagtgtgtcctggaagact 1062
 QY 301 ggaagcactgtctgtgtgacttacttcccatcacatgactgctgaggggtggaatt 360
 DB 1063 ggaagcactgtctgtgtgacttacttcccatcacatgactgctgaggggtggaatt 1122
 QY 361 caaggtcagcaaggaataaacaagttaaaggagcactcccaactgaccggggaat 420
 DB 1123 caaggtcagcaaggaataaacaagttaaaggagcactcccaactgaccggggaat 1182
 QY 421 agcaccctagccctacaaagagtagttctctcaaaactacatgaaacccctcgatcc 480

```

ID  ABA45822 standard; DNA; 1894 BP.
XX
XX  ABA45822;
AC
XX
XX  01-FEB-2002 (first entry)
DT
XX
XX  Human breast cell single exon nucleic acid probe #4517.
DE
XX
XX  Human; microarray; single exon probe; gene expression; breast;
KW  disease; cancer; ss.
XX
XX  Homo sapiens.
XX
XX  WO200157271-A2.
PN
XX
XX  09-AUG-2001.
PD
XX
XX  30-JAN-2001; 2001WO-US00662.
PF
XX
XX  04-FEB-2000; 2000US-0180312.
PR  26-MAY-2000; 2000US-0207456.
PR  30-JUN-2000; 2000US-0608408.
PR  03-AUG-2000; 2000US-0632366.
PR  21-SEP-2000; 2000US-0234687.
PR  27-SEP-2000; 2000US-0236359.
PR  04-OCT-2000; 2000GB-0024263.
PA
XX  (MOLE-) MOLECULAR DYNAMICS INC.
PI
XX  Penn SG, Hanzel DK, Chen W, Rank DR;
XX  WPI; 2001-496933/54.
DR
XX
XX  New spatially-addressable set of single exon nucleic acid probes,
PT  useful for measuring gene expression in sample derived from human
PT  breast, comprises number of single exon nucleic acid probes
XX
XX
XX  Claim 1; SEQ ID NO 4517; 327bp + sequence listing; English.
PS
XX
XX  The invention relates to a spatially-addressable set of single exon
CC  nucleic acid probes for measuring gene expression in a sample derived
CC  from human breast and BT 474 cells. The method involves contacting
CC  the probes with a collection of detectably labelled nucleic acids
CC  derived from mRNA of human breast, and then measuring the label
CC  bound to each probe of the microarray. The probes are useful for
CC  verifying the expression of regions of genomic DNA predicted to
CC  encode proteins. They are useful for gene discovery, and for
CC  determining predisposition and/or prognosing breast disease. Gene
CC  expression analysis is useful for assessing the toxicity of chemical
CC  agents on cells. The microarray of this invention presents a far greater
CC  diversity of probes for measuring gene expression, with far less bias
CC  than expressed sequence tag microarrays. The method is suitable for
CC  rapid production of functional information from genomic sequence. The
CC  present sequence is a single exon nucleic acid probe of the invention.
CC  Note: The sequence data for this patent did not form part of the
CC  printed specification, but was obtained in electronic format directly
XX  from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX  Sequence 1894 BP: 557 A; 500 C; 361 G; 476 T; 0 other;
SO

Query Match 76.4%; Score 1131.6; DB 22; Length 1894;
Best Local Similarity 92.4%; Pred. No. 0;
Matches 1218; Conservative 0; Mismatches 84; Indels 16; Gaps
2;

07 164 tatctaaagggaaactccacccttcaactgcccacacaccattatgccccgcgaactgcataact 223
db 1 tatctaaaggaaccccccacttcaactgcccacacaccattatgccccacacactgcataact 60

```



```
QY 284 gtgtctcgtgaggactgtagagccactgtctgtggaacttactcaaccataccagatgt 343
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 gtgtccctgtaggacttggg-----ggactcaactcaatcactacagatgt 167
QY 344 ctgattggtggtggaattccaagttcagagcaagagaanaaagtaaggaagcaatcc 403
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 168 ctgattggtggtggaattccaagttcagagcaagagaanaaagtaaggaagtaatcc 227
QY 404 aactgagccgggagacatagacccttagccctcacaagaagactagtctctcaaaaact 463
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 228 aactgagctgggtacatagacccttagccctcacaagaagactagtctctcaaaaact 287
QY 464 atgaaacctcgttacacactcgtctggtgagcctatttaatacaccactcactcgtgc 523
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 288 atgaaacctcactacactcgtctggtgagcctatttaatacaccactcactcgtgc 347
QY 524 tccattgaggtctcagcccaaacctcactactcgtgtgagatgtgcctccctcacttca 583
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 348 tccattgaggtctcagcccaaacctcactactcgtgtgagatgtgcctccctcacttca 407
QY 584 ggcatacatctcaatccctgtctcgaacaatggaacaacttcagacagacaagaataaaca 643
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 408 ggcatacatctcaatccctgtctcgaacaatggaacaacttcagacagacaagaataaaca 467
QY 644 ccaacttcgttttagtaggaacctctgtttcccaatctggaataaccatacctcaaac 703
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 468 ccaacttcgttttagtaggaacctctgtttcccaatctggaataaccatacctcaaac 524
QY 704 tcaacctgtgtaaaatttagaataactataacacaaacagctcccaatgcatacagttgg 763
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 525 tcaacctgtgtaaaatttagaataactataacacaaacagctcccaatgcatacagttgg 584
QY 764 taacaacctccacaagaatgtctgacctaccctcagaagaatattttgtctgtgtgtaact 823
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 585 taactcctccacaagaatgtctgacctaccctcagaagaatattttgtctgtgtgtaact 644
QY 824 cagactatcatgtttggaatggtctcagaatcatalgtgtctcctcactcacttaagtc 883
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 645 cagactatcatgtttggaatggtctcagaatcatalgtgtgtctcctcactcacttaagtc 704
QY 884 cccctttagaccatctacacacgaagaattatatacaatctgtctgtaacgaagccca 943
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 705 cccctttagaccatctacacacgaagaattatatacaatctgtctgtaacgaagccca 764
QY 944 acaaaagtagaccatctcctcttctgttatacagagagagtgctagggagtagta 1003
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 765 acaaaagtagaccatctcctcttctgttatacagagagagtgctagggagtagta 824
QY 1004 ctgcatgtgcaatcatalcacaacctcactcagttcactaactcaactaactcaagaataa 1063
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 825 ctgcatgtgcaatcatalcacaacctcactcagttcactaactcaactaactcaagaataa 884
QY 1064 atgtgtacatggaacagtgtaactgactcctcctgtgtacacttgcagaatcaactcc 1123
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 885 atgtgtacatggaacagtgtaactgactcctcctgtgtacacttgcagaatcaactcc 944
QY 1124 tagcagtagtagtccctcaaatcgaagagcttagaactgtgtaacggccaagaagggg 1183
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 945 tagcagtagtagtccctcaaatcgaagagcttagaactgtgtaacggccaagaagggg 1004
QY 1184 gaacctgttatttttagagaagaacgctgtattatgttaactcaatcagaatttca 1243
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1005 gaacctgttatttttagagaagaacgctgtattatgttaactcaatcagaatttca 1064
QY 1244 ctgagaagaagttaagaataatcagagatcgaaatcagatgtaagcagagagatcacaaca 1303
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1065 ccgagaagaagttaagaataatcagagatcgaaatcagatgtaagcagagagatcacaaca 1124
QY 1304 ccgaacgcttggtggtcctcctcagccaatggaatggtcctggtcctcctttagagactc 1363
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1125 ctggaaccttggtggtcctcctcagccaatggaatggtcctggtcctcctttagagactc 1184
QY 1364 tagcagctcctaataatgttactcctctttagaacctgtatcttaacctctgttgaat 1423
```

```
Db 1185 tagcagctataataatgttactcctctttagaacctgtatcttaactcctgttgaat 1244
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1424 ttgtctctccagaattggaagctgtgaaagctacagatgtgtcttaacatggaaccaca 1481
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1245 ttgtctctccagaattggaagctgtgaaagctacagatgtgtcttaacatggaaccaca 1302

RESULT 15
ABA56337
ID ABA56337 standard; DNA; 1894 BP.
XX
AC ABA56337;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #4642.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PE 30-JAN-2001; 2001MO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236559.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX
PS analyzing gene expression in human foetal liver -
XX
XX
PS Claim 1; SEQ ID NO 4642; 639pp + sequence listing; English.
XX
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC foetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 1894 BP; 557 A; 500 C; 361 G; 476 T; 0 other;

Query Match 76.4%; Score 1131.6; DB 22; Length 1894;
Best Local Similarity 92.4%; Pred. No. 0;
Matches 1218; Conservative 0; Mismatches 84; Indels 16; Gaps 2;
```

```
QY 164 tatctaaaggaacatccactcactgcccacacacatagccccgaactgtataact 223
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 tatctaaaggaaccccccactcactgcccacacacatagccccgaactgtataact 60
QY 224 ctggcactcttggatgcatgcaataactatattatggaaggaagaatgataactcta 283
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 ctggcactcttggatgcatgcaataactatattatggaaggaagaatgataactcca 120
QY 284 gtgtcctgtagagacttgagccactgtctgtgacttactcaaccataccagatgt 343
```

Db 121 gtgtctcgtgaggaacttga-----ggactcaactctctctacacagatgt 167
Qy 344 cggatcgggctggaattcgaagtcagggcaagaagaacaaacgaagtaaggagcaatctccc 403
Db 168 cggatcgggctggaattcgaagtcagggcaagaagaagaacaaacgaagtaaggagcaatctccc 227
Qy 404 aactgaaccggggaacatagcaacccttagccctacaaagaagtaagtctctcaaaactac 463
Db 228 aactgaaccggggaacatagcaacccttagccctacaaagaagtaagtctctcaaaactac 287
Qy 464 atgaaaccctcgttccctactcgtcgtgtgagcctatttaaccactcactcggc 523
Db 288 atgaaaccctcgttccctactcgtcgtgtgagcctatttaaccactcactcggc 347
Qy 524 tccatgaagtcacagcccaaacctactaactgtgtgga tgtgcctccctgcacttca 583
Db 348 tccatgaagtcacagcccaaacctactaactgtgtgga tgtgcctccctgcacttca 407
Qy 584 ggcataacattcaatccctgttccctgaaacaa tggaaacaaactcagcacagaataaaca 643
Db 408 ggcataacattcaatccctgttccctgaaacaa tggaaacaaactcagcacagaataaaca 467
Qy 644 ccaacttcggttttagtaagacctctgtttccaatctggaataaaccatacctaacc 703
Db 468 ccaacttcggttttagtaagacctctgtttccaatctggaataaaccatacctaacc 524
Qy 704 tcaactgtgtaaatlttaagaataactatagaacaaacagccccaatgacatcagttggg 763
Db 525 tcaactgtgtaaatlttaagaataactatagaacaaacagccccaatgacatcagttggg 584
Qy 764 taacaactcccaacgaatagtcgcctacccctcagaataattttgtcgtgtaacct 823
Db 585 taacaactcccaacgaatagtcgcctacccctcagaataattttgtcgtgtaacct 644
Qy 824 cagcctatcatgtgttgaaatggtctctcagaatactatgtgctcctctcatctagtg 883
Db 645 cagcctatcatgtgttgaaatggtctctcagaatactatgtgctcctctcatctagtg 704
Qy 884 cccctaagcaactcaacatgaaagatltatcaaatcatgtcgtacaaagcccca 943
Db 705 cccctaagcaactcaacatgaaagatltatcaaatcatgtcgtacaaagcccca 764
Qy 944 acaaaaagataccatctctcttctgtatcaagagcagagtgctagggcagactagta 1003
Db 765 acaaaaagataccatctctcttctgtatcaagagcagagtgctagggcagactagta 824
Qy 1004 ctggcatgtgcgtatcaaaccttactcagttctactacaactatctcaagaataa 1063
Db 825 ctggcatgtgcgtatcaaaccttactcagttctactacaactatctcaagaataa 884
Qy 1064 atggtgaactggaacaggtcacatgtaacctcctgtcaccttgcagaatcaacttaactccc 1123
Db 885 atggtgaactggaacaggtcgctgtataccctgtcaccttgcagaatcaacttaactccc 944
Qy 1124 tagcagcagtagtctctcaaaatcgaaagacttagacttgcataacggcaaaagaggg 1183
Db 945 tagcagcagtagtctctcaaaatcgaaagacttagacttgcataacggcaaaagaggg 1004
Qy 1184 gaaactgtttatttttagaagaagaactgtgtattatgttaataatccagaattgtca 1243
Db 1005 gaaactgtttatttttagaagaagaactgtgtattatgttaataatccagaattgtca 1064
Qy 1244 ctgagaagaatcaagaatctgaaatcgaaataatgttagagcagagaggtctcaaaaca 1303
Db 1065 ctgagaagaatcaagaatctgaaatcgaaataatgttagagcagagaggtctcaaaaca 1124
Qy 1304 ccgaaacgtctgggctcctcctcagccaatggaatgcccctgggtctccctcttagagactc 1363
Db 1125 ctggacccctgggctcctcctcagccaatggaatgcccctgggtctccctcttagagactc 1184
Qy 1364 tagcagcctaaatattgttaactcctcttggacccctgtatcttaactccttgttaagt 1423

Db 1185 tagcagcctaataatgttaactcctccttggacccctgtatctttaaactccttgttaagt 1244
Qy 1424 tgtctctccagaattggaactgttaagctacagaatggtcttaacaatatggaaccca 1481
Db 1245 tgtctctccagaattggaactgttaagctacagaatggtcttaacaatatggaaccca 1302

Search completed: June 20, 2002, 09:43:22
Job time: 8532 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 09:43:49 ; Search time 4429.18 Seconds
(without alignments)
6997.277 Million cell updates/sec

Title: US-09-319-156A-9
Perfect score: 1481
Sequence: 1 atggccctccctatacatc.....gtctacaatcggaaccaca 1481

Scoring table: IDENTITY-NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues 3595312
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_pro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
------------	-------------	-------	--------	----	----	-------------

1	1481	100.0	1481	6	AX001027	AX001027 Sequence
2	1481	100.0	1932	14	AF127228	AF127228 Multiple
3	1468.2	99.1	162579	9	AL390039	AL390039 Human DNA
4	1461.8	98.7	1629	6	AX036992	AX036992 Sequence
5	1461.8	98.7	1629	14	AF331500	AF331500 Multiple
6	1461.8	98.7	172918	2	AC023366	AC023366 Homo sapi
7	1335.4	90.2	2930	6	AX355872	AX355872 Sequence
8	1335.4	90.2	2930	6	AF208161	AF208161 Homo sapi
9	1335.4	90.2	2946	6	AR177269	AR177269 Sequence
10	1333.8	90.1	2055	6	AX007999	AX007999 Sequence
11	1333.8	90.1	2539	6	AX007978	AX007978 Sequence
12	1333.8	90.1	10499	6	AX007980	AX007980 Sequence
13	1333.8	90.1	56093	6	AX329572	AX329572 Sequence
14	1333.8	90.1	56093	6	HSAC000064	AC000064 Human BAC
15	1333.8	90.1	149194	9	AC007566	AC007566 Homo sapi
16	1332.2	90.0	1617	9	AF156963	AF156963 Homo sapi
17	1330.6	89.8	2781	6	AX092223	AX092223 Sequence
18	1330.6	89.8	2781	6	AF072506	AF072506 Homo sapi
19	1324.2	89.4	7582	6	AX000966	AX000966 Sequence
20	1324.2	89.4	7582	6	AX027480	AX027480 Sequence
21	1318	89.0	15344	2	AC040948	AC040948 Homo sapi
22	1318	89.0	205035	9	CNS00005	AL049870 Human chr
23	1306.6	88.2	2782	6	AX000962	AX000962 Sequence
24	1306.6	88.2	2782	6	AX027476	AX027476 Sequence
25	1306.6	88.2	2782	6	AX133396	AX133396 Sequence
26	1297.4	87.6	161571	2	AC092708	AC092708 Homo sapi
27	1297.4	87.6	166700	9	AC013759	AC013759 Homo sapi
28	1297.4	87.6	186911	2	AP001488	AP001488 Homo sapi
29	1294	87.4	172721	2	AC008121	AC008121 Homo sapi
30	1285	87.4	188919	2	AC009727	AC009727 Homo sapi
31	1285	86.8	162522	9	AC093839	AC093839 Homo sapi
32	1282	86.6	46340	6	AX008043	AX008043 Sequence
33	1282	86.6	177163	9	HS134E15	AL022067 Human DNA
34	1278.8	86.3	1799	6	AX007982	AX007982 Sequence
35	1276.2	86.2	188038	2	AC092510	AC092510 Papio cyn
36	1271.2	85.8	178203	9	AC008780	AC008780 Homo sapi
37	1261	85.1	158033	9	AC018926	AC018926 Homo sapi
38	1258.8	85.0	169462	2	AC024033	AC024033 Homo sapi
39	1256.2	84.8	182178	9	AC074198	AC074198 Homo sapi
40	1255.8	84.8	183680	9	AC098859	AC098859 Homo sapi
41	1251.6	84.5	113253	2	AC092356	AC092356 Homo sapi
42	1247.8	84.3	186723	2	AC027752	AC027752 Homo sapi
43	1242.2	83.9	183267	2	AC068447	AC068447 Homo sapi
44	1239	83.7	2784	6	AX007981	AX007981 Sequence
45	1239	83.7	135734	2	AC002346	AC002346 Homo sapi

ALIGNMENTS

RESULT	1	1481 bp	DNA	Linear	PAT	10-MAR-2000
AX001027	AX001027	Sequence 9 from Patent WO9902666.				
LOCUS	AX001027	1481 bp	DNA	Linear	PAT	10-MAR-2000
DEFINITION	AX001027	Sequence 9 from Patent WO9902666.				
ACCESSION	AX001027					
VERSION	AX001027.1	GI:7241264				
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
source						
BASE COUNT						
ORIGIN						

1.1481
/organism="unidentified"
/db_xref="taxon:32644"
412 a 410 c 261 g 398 t

ASSOCIATED WITH MULTIPLE SCLEROSIS AND RHEUMATOID ARTHRITIS, FOR
DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES
Patent: WO 9902666-A 9 21-JAN-1999;
BIO MERIEUX (FR); OTT CATHERINE (FR)
location/Qualifiers

IMPORTANT: This sequence is not the entire insert of clone RPJ3-383K5. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RPJ3-383K5 is at 1 in this sequence. The true left end of clone RPJ3-383K5 is at 162480 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. RPJ3-383K5 is from the library RPJ3-13.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

FEATURES

```
Source
1. 162579
/organism="Homo sapiens"
/dbxref="taxon:9606"
/chromosome="X"
/map="q22.1-24"
/clone="RPJ3-383K5"
/clone.lib="RPJ3-13.2"
complement(1..232)
/note="match: GSS: Em:A0036363"
16..354
/note="match: GSS: Em:A0504848"
475..644
/note="AluSg repeat: matches 134..303 of consensus"
742..919
/note="MLT1B repeat: matches 207..388 of consensus"
1112..1142
/note="L1MC repeat: matches 327..356 of consensus"
1143..1456
/note="AluSx repeat: matches 1..312 of consensus"
1457..2938
/note="L1MC repeat: matches 356..1897 of consensus"
2922..3739
/note="L1PA2 repeat: matches 5329..6146 of consensus"
3844..4087
/note="L1M4 repeat: matches 1..251 of consensus"
4092..4231
/note="L1M4 repeat: matches 612..756 of consensus"
4373..4559
/note="L1MC repeat: matches 971..1152 of consensus"
4586..4619
/note="17 copies 2 mer aa 100% conserved"
4623..4964
/note="MLT1A1 repeat: matches 3..333 of consensus"
4989..5149
/note="L1M8 repeat: matches 6120..6288 of consensus"
5152..5602
/note="L1PA15 repeat: matches 5325..5779 of consensus"
5606..5916
/note="AluSg repeat: matches 1..313 of consensus"
5917..6014
/note="49 copies 2 mer ga 75% conserved"
6018..6409
/note="L1PA15 repeat: matches 5769..6157 of consensus"
6566..6750
/note="L1M4 repeat: matches 2857..3036 of consensus"
6751..7063
/note="AluJo repeat: matches 1..307 of consensus"
7064..8097
/note="L1M4 repeat: matches 3036..4105 of consensus"
8099..8419
/note="AluSx repeat: matches 1..306 of consensus"
8420..8487
/note="L1MC/D repeat: matches 5582..5648 of consensus"
10039..10396
/note="THE1B repeat: matches 1..364 of consensus"
10519..10807
/note="AluSg repeat: matches 3..297 of consensus"
11343..11416
```

```
repeat_region
11448..11983
/note="37 copies 2 mer tt 67% conserved"
repeat_region
12030..12337
/note="L1MD3 repeat: matches 7162..7739 of consensus"
repeat_region
12995..13077
/note="AluSx repeat: matches 1..309 of consensus"
repeat_region
13110..13222
/note="MIR repeat: matches 103..186 of consensus"
repeat_region
13228..13518
/note="FLAMC repeat: matches 2..124 of consensus"
repeat_region
13519..13853
/note="MER7A repeat: matches 1..346 of consensus"
repeat_region
13854..14039
/note="L1MC repeat: matches 1951..2131 of consensus"
repeat_region
14040..14254
/note="MER20 repeat: matches 1..218 of consensus"
repeat_region
14255..14807
/note="L1MC repeat: matches 1686..1951 of consensus"
repeat_region
14808..15105
/note="AluY repeat: matches 1..311 of consensus"
repeat_region
15106..15492
/note="L1MC repeat: matches 1309..1686 of consensus"
repeat_region
15496..15822
/note="L1M5 repeat: matches 5823..6167 of consensus"
repeat_region
15824..15920
/note="L1MC repeat: matches 1044..1140 of consensus"
repeat_region
15885..16086
/note="match: STS: Em:G34444 Em:L14255"
15920..16082
/note="match: GSS: Em:A0769824"
complement(15927..16073)
/note="match: GSS: Em:A0054794 Em:AQ823559"
16067..16294
/note="match: GSS: Em:AQ457003"
16923..16998
/note="38 copies 2 mer at 71% conserved"
17107..17166
/note="L2 repeat: matches 2688..2750 of consensus"
17175..17755
/note="L2 repeat: matches 1244..1898 of consensus"
17839..18106
/note="MER70A repeat: matches 148..418 of consensus"
18213..18318
/note="L2 repeat: matches 1128..1247 of consensus"
18706..18811
/note="L2 repeat: matches 2631..2750 of consensus"
18812..19057
/note="AluSx repeat: matches 1..297 of consensus"
19058..19492
/note="L2 repeat: matches 2213..2647 of consensus"
complement(19352..19843)
/note="match: GSS: Em:AQ340313"
19556..19872
/note="UTR17 repeat: matches 1..326 of consensus"
19873..21594
/note="HERV17 repeat: matches 6812..8523 of consensus"
21587..22280
/note="HERV17 repeat: matches 4301..4980 of consensus"
22298..22742
/note="L1M5 repeat: matches 5623..6135 of consensus"
22743..23048
/note="AluY repeat: matches 1..306 of consensus"
23049..23286
/note="L1M5 repeat: matches 5403..5623 of consensus"
23285..23361
/note="L1M4 repeat: matches 5006..5082 of consensus"
23364..23429
/note="L1M4 repeat: matches 4019..4084 of consensus"
23429..24305
/note="L1 repeat: matches 3097..4043 of consensus"
24306..24593
/note="AluJo repeat: matches 1..293 of consensus"
```


TITLE
JOURNAL
COMMENT

Boukhaltier, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
Chapel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
Dearellano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M.,
Fenster, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D.,
Galligan, J., Gardyna, S., Glade, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Hearford, A., Horton, L.,
Klein, J., Landers, T., Largocque, K., Lehoczy, J., Levine, R.,
Lieu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N., McCarthy, M.,
McMan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J.,
Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M.,
Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C.,
Riley, R., Rogov, P., Rochman, D., Roy, A., Santos, R., Schauer, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Sutramanlan, A., Talamas, J., Testaye, S., Theodore, J., Titrrell, A.,
Travers, M., Trigllo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
Wu, X., Wyman, D., Ye, W. J., Young, G., Zalnoun, J., Zimmer, A. and
Zody, M.

Direct Submission
Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 2, 2000 this sequence version replaced gi:6970363.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L3492
Center clone name: 15_N_10

----- Summary Statistics
Sequencing vector: M13: M7815: 100% of reads
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.960731
Consensus quality: 142569 bases at least Q40
Consensus quality: 156863 bases at least Q30
Consensus quality: 164107 bases at least Q20
Insert size: 178000: agarose-fp
Insert size: 169718: sum-of-contigs
Quality coverage: 2.9 in Q20 bases; agarose-fp
Quality coverage: 3.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 33 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1266: contig of 1266 bp in length
* 1267 1366: gap of 100 bp
* 1367 2686: contig of 1320 bp in length
* 2687 2786: gap of 100 bp
* 2787 4288: contig of 1502 bp in length
* 4289 4388: gap of 100 bp
* 4389 5572: contig of 1184 bp in length
* 5573 5672: gap of 100 bp
* 5673 6812: contig of 1140 bp in length
* 6813 6912: gap of 100 bp
* 6913 7404: contig of 492 bp in length
* 7405 7504: gap of 100 bp
* 7505 8913: contig of 1411 bp in length
* 8916 9015: gap of 100 bp
* 9016 10200: contig of 1185 bp in length
* 10201 10300: gap of 100 bp
* 10301 12003: contig of 1703 bp in length
* 12004 12103: gap of 100 bp
* 12104 14038: contig of 1935 bp in length

* 14039 14138: gap of 100 bp
* 14139 16226: contig of 2088 bp in length
* 16227 16326: gap of 100 bp
* 16327 19115: contig of 2789 bp in length
* 19116 19215: gap of 100 bp
* 19216 21492: contig of 2277 bp in length
* 21493 21592: gap of 100 bp
* 21593 24794: contig of 3202 bp in length
* 24795 24894: gap of 100 bp
* 24895 27624: contig of 2730 bp in length
* 27625 27724: gap of 100 bp
* 27725 30780: contig of 3056 bp in length
* 30781 30880: gap of 100 bp
* 30881 34253: contig of 3373 bp in length
* 34254 34353: gap of 100 bp
* 34354 38260: contig of 3907 bp in length
* 38261 38360: gap of 100 bp
* 38361 42757: contig of 4397 bp in length
* 42758 42857: gap of 100 bp
* 42858 46232: contig of 3375 bp in length
* 46233 46332: gap of 100 bp
* 46333 51049: contig of 4717 bp in length
* 51050 51149: gap of 100 bp
* 51150 53373: contig of 2224 bp in length
* 53374 53473: gap of 100 bp
* 53474 58170: contig of 4697 bp in length
* 58171 58270: gap of 100 bp
* 58271 63398: contig of 5128 bp in length
* 63399 63498: gap of 100 bp
* 63499 69283: contig of 5785 bp in length
* 69284 69383: gap of 100 bp
* 69384 75177: contig of 5794 bp in length
* 75178 75277: gap of 100 bp
* 75278 80817: contig of 5540 bp in length
* 80818 80917: gap of 100 bp
* 80918 89553: contig of 8636 bp in length
* 89554 89653: gap of 100 bp
* 89654 101099: contig of 11446 bp in length
* 101100 101199: gap of 100 bp
* 101200 113135: contig of 11936 bp in length
* 113136 113235: gap of 100 bp
* 113236 132520: contig of 19285 bp in length
* 132521 132620: gap of 100 bp
* 132621 150646: contig of 18026 bp in length
* 150647 150746: gap of 100 bp
* 150747 172918: contig of 22172 bp in length.

FEATURES
source
1. 172918
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-15N10"
/clone_1lb="RPC1-11 Human Male BAC"
1. 1266
/note="assembly-fragment"
misc_feature
1367. 2686
/note="assembly-fragment"
misc_feature
2787. 4288
/note="assembly-fragment"
misc_feature
4389. 5572
/note="assembly-fragment"
misc_feature
5673. 6812
/note="assembly-fragment"
misc_feature
6913. 7404
/note="assembly-fragment
clone_end.r7
vector_slide:right"
7505. 8913
/note="assembly-fragment"
misc_feature
9016. 10200
/note="assembly-fragment"
misc_feature
10301. 12003
/note="assembly-fragment"
12104. 14038
misc_feature

```
/note="assembly_fragment"  
misc_feature 14139..16226  
/note="assembly_fragment"  
misc_feature 16327..19115  
/note="assembly_fragment"  
misc_feature 19216..21492  
/note="assembly_fragment"  
misc_feature 21593..24794  
/note="assembly_fragment"  
misc_feature 24895..27624  
/note="assembly_fragment"  
misc_feature 27725..30780  
/note="assembly_fragment"  
misc_feature 30881..34253  
/note="assembly_fragment"  
misc_feature 34354..38260  
/note="assembly_fragment"  
misc_feature 38361..42757  
/note="assembly_fragment"  
misc_feature 42858..46232  
/note="assembly_fragment"  
misc_feature 46333..51049  
/note="assembly_fragment"  
misc_feature 51150..53373  
/note="assembly_fragment"  
/note="assembly_fragment"  
vector_end:SP6  
clone_end:right"  
53474..58170  
/note="assembly_fragment"  
misc_feature 58271..63398  
/note="assembly_fragment"  
misc_feature 63499..69283  
/note="assembly_fragment"  
misc_feature 69384..75177  
/note="assembly_fragment"  
misc_feature 75278..80817  
/note="assembly_fragment"  
  
Query Match 98.7%; Score 1461.8; DB 2; Length 172918;  
Best Local Similarity 99.2%; Pred. No. 0;  
Matches 1469; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
  
QY 1 atggccctccctacatcattctctcttactgtctcttacccttcgctctcat 60  
Db 103234 ATGGCCCTCCCTATCATATTTTCTTTACTGTCTCTCACCCCTTGCCTCACT 103293  
  
QY 61 gcaacccctcatgtctgtgtaacacacagtagtcccttaccgaagattctatgaaga 120  
Db 103294 GCACCGCTTCCATGCTGTGTAACAACAGTAGCTCCCTTACCAAGAGTTTCTATAAAGA 103353  
  
QY 121 acgagctctcctggaatgatgatgcccacatagagagttatctcaagggaacacc 180  
Db 103354 ACGCGGCTTCCGGAATATGTGATGCCCATCATATAGAGTTTATCTAAGGAAATCC 103413  
  
QY 181 acccttactgcccacaccacatattgcccgcgaactgctataactctgcactcttgcatt 240  
Db 103414 ACCTTACTGCCCCACACCACTATGCCCCGCAACTGCTAATACTTGCCACTTTTGCAATG 103473  
  
QY 241 catgnaaatcatcatatttgacagaggaataatgatataatccagtgtctctggaagactt 300  
Db 103474 CATGAAATACTCATATTATGGACAGGAAATGATTAATCCATGTTGCTTGAGAGACTT 103533  
  
QY 301 ggaagcactgtctgttgaacttaccacccataccagatgtctgattgagggtgaagt 360  
Db 103534 GGAAGCACTGTCTGTGGACTTACTTACCCATACAGTATGTCTGATGGGGGTGAATTT 103593  
  
QY 361 caaggtcagggcaagaagaacaagtaaaaggaagcaatctcccaactgaccggggagact 420  
Db 103594 CAAGGTCAAGGCAAGAAAGAAACAGTAAGAAGCAATCTCCCACTGACCCGGGAGACT 103653  
  
QY 421 agcacccttagccctacaaagagatagtctctcaaaactacatgaagaccctcgctac 480  
Db 103654 AGCACCCTTAGCCCTACAAAGAGACTAGTTCTCTCAAAACTACATGAAGAACCTCGGTACC 103713
```

```
QY 481 catacgccctggtagccatttaatacacaccctcactcgctccatgaggtcgaacc 540  
Db 103714 CATACTGCGCTGGTAGGCTTTTATATACACCTCACTCGGCTTCATGAGAGTCTCAGCC 103773  
  
QY 541 caaaacccctactaacctgttgaatgctcccccctgcacttcgaagccatattcaatc 600  
Db 103774 CAAAACCTTACTAACTGTTGGATGTGCTCCCGCTGCACCTTCAAGGCCATATATTCAATC 103833  
  
QY 601 cctgttctgaaacatggaacaacttcgaacagaataaaccacacttccgttttagta 660  
Db 103834 CCTGTCTGGAACATGGAACAACTTCAGACAGAAATTAACACCACTTCGTTTACTA 103893  
  
QY 661 ggaaccttcttcccaactctgaaataacccataccctcaaacctcactgtgtataatt 720  
Db 103894 GGACCTCTGTTTCCAACTTGGAATTAACCATACCTCAAACTCAGCTGTGTAAATTT 103953  
  
QY 721 agcaatcatagacacacacagactcccaatgacatgaagtggtaacacctccacaga 780  
Db 103954 AGCAATACTATAGACACAAACAGCTCCCAATGACATCAGTGGGTAACTCCACAGCA 104013  
  
QY 781 atagtctccctacccctcaggaatatatttctgtgtgttacctgaagctcatcatgtttg 840  
Db 104014 ATAGCTGCTCAACCTCGAATATATTTTGTGTGTGTGCTACCTCAGCTATCATTTGTTG 104073  
  
QY 841 aatggctctcagaatctatgtctctcctccatctcttagtggccctatgacacatc 900  
Db 104074 AATGGCTTTTCAAGANTCTATGTGCTCTCTCATATTTTAGTGCCCTCATAGCAATCTAC 104133  
  
QY 901 actgacacagaattatataacatcgtctgtaactaaagcccaacaaagaagtaaccatt 960  
Db 104134 ACTGAACAAGATTTATACAAATCATGTGCTACTTAAGCCCAACAAAGAGTACCATT 104193  
  
QY 961 ctctcttctgtatcagagcagagagtgctgtagcagactcaggtactcgtgaattggcagatc 1020  
Db 104194 CTTCCTTTGTTATCAAGAGAGAGAGTGTAGCAACATGAGTACGATTTGGCATGTATC 104253  
  
QY 1021 acaacctactcaatcttactacacaaactatctcaagaataaatggtgacatggaacag 1080  
Db 104254 ACAACCTTACTCACTGCTTACTACAACTATCTCAAGAAATTAAGTGACATGGAACAG 104313  
  
QY 1081 gtcactgacccctggtgcacacttgcagaatcaacttaactccctcagcagtagtctct 1140  
Db 104314 GTCACTGACTCCCTGTGCTACCTTGCAGAAATCACTTAACCTCCCTGACAGCATGACTCCTT 104373  
  
QY 1141 caaaatcgaagagctttagactgttaacccgcaaaagaaggaggaacctgtttatttta 1200  
Db 104374 CAAAATCGAAGAGCTTTAGACTTGTAAACCGCAAAAGAGGGGGAACCTGTTATTTTAA 104433  
  
QY 1201 ggaagaagaagctgttattatgttaatcaatcagaatgttctactgagaagaattaaagaa 1260  
Db 104434 GGAGAAGAATGCTGTTATTAATTAATCAATGCAAAATGTGCAGAGAAAGTTAAAGA 104493  
  
QY 1261 atcgagatcgaaatacaatgtgagcagagagagcttcaaaacacggaagctgggctc 1320  
Db 104494 ATTCAGATTCGAATCAATGTGAGACGAGAGGCTTCAAAACCCGAACACAGCTGGGCGCTC 104553  
  
QY 1321 ctcaagcaatgatatgacctgggttctccctcttaagaaacctcagcagcttaatatgt 1380  
Db 104554 CTCAGCCCAATGGAATCCCTGGGTTCTCCCTTTTAGGACCTTGACAGCTTAATATTG 104613  
  
QY 1381 ttaactccctcttggagccctgtatcttaaccctccctgttlaagtctgtcctccagaatt 1440  
Db 104614 TTACTCTCTTTTGGACCCGTATCTTTAACTCTTGTAAATTTGTCTTCCGAATTT 104673  
  
QY 1441 gaagctgtlaaagctacagatggtcttaacaatggaacccca 1481  
Db 104674 GAAGCTGTAAAGCTCAAAATGCTTCTTCAAAATGAGGCCCA 104714  
  
RESULT 7  
AA355872 2930 bp DNA linear PAT 06-FEB-2002  
LOCUS
```


Accession	Version	Keywords	Source	Organism	Reference Authors	Title	Journal	Medline Reference	Authors
AF208161	1	GI:6760400	human.	human.					
				human.					
				Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
				Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
				1 (bases 1 to 2930)					
				Mi.S., Lee.X., Li.X., Veldman,G.M., Finnerty,H., Racie,L.,					
				Lavallee,E., Tang,X.Y., Edouard,P., Howes,S., Keith,J.C. Jr. and					
				McCooy,J.M.					
				Syncytin is a captive retroviral envelope protein involved in human					
				placental morphogenesis					
				Nature 403 (6771), 785-789 (2000)					
				2 (bases 1 to 2930)					
				Sha,M., Lee,X., Li,X., Veldman,G.M., Finnerty,H., Racie,L.,					
				Lavallee,E., Tang,X., Edouard,P., Howes,S., Keith,J.C. Jr. and					
				McCooy,J.M.					
				Direct Submission					
				Submitted (26-NOV-1999) Genetics Institute, 87 Cambridge Park					
				Drive, Cambridge, MA 02140, USA					
				Location/Qualifiers					
				1..2930					
				/organism="Homo sapiens"					
				/db_xref="taxon:9606"					
				/clone="AJ172-2s"					
				/rname="type="testis"					
				1..2930					
				/rpe_family="human endogenous retrovirus HERV-W"					
				1..929					
				930..2546					
				/note="envelope protein"					
				/codon_start=1					
				/product="syncytin precursor"					
				/protein_id="AA28334.1"					
				/db_xref="GI:6760401"					
				/translation="MALPRLHFLFTVLLPSPFLTAPPCRCMTSSPYOFLWMRPP					
				GNIVAPRSLSKGTPTETATHRMRCYSHATCMANHTYMGKMINSCPGGLGV					
				TWCWYFTQMSGGGVODAREKHREYISOLTPVHGSSPYKGLDSKLTHTLT					
				HTRLVSLTHTTGLTHESAVQNPWCICLPLNRPVYSIPVPEOMNPFTEINTTSV					
				LGVGLVSLTHTTGLTHESAVQNPWCICLPLNRPVYSIPVPEOMNPFTEINTTSV					
				YRLGNSSESKCFSLFPPPTITTEDDLNIVISKFRNKRPVLPVIGAGVGAUG					
				TGIGITTSIQFYKRLSQELNGDMERVADSLVTLQDLSNLAVALONRRALDLTAE					
				RGCGCLFGEECCYVNOGIVTEKVEIRDRIOREAEELNTEGPGMLLSQWMPILP					
				FLVGLAAILLLLEGPCLFNLVNFVSRIEAVYLQMEPKMQSKTKIYRRLDRPASP					
				RSDVNDIKGTPPEPISAAQPLLRNSAGSS"					
				930..989					
				990..2543					
				/product="syncytin"					
				1869..1880					
				/note="furin; cleavage site"					
				/evidence=not_experimental					
				2043..2120					
				/note="Region: immunosuppressive region"					
				2235..2357					
				/note="transmembrane-region site"					
				2547..2930					
				BASE COUNT					
				842 a 800 c 571 g 717 t					

OY	121	acgagcttcctggaatatgtgtgccccatcatagaagttatctcaaggaaaccc	180
Db	1050	ATGCAAGGTCCGGAAATATTTGATGCCCCACTATATGAGAGCTTCTTCAAGGAACCCCC	1109
OY	181	acctcaactgcccacaccatcatgcccgaaactgtataactctgcacactcttgcatg	240
Db	1110	ACCTTCACTGCCCCACACCCTATGCCCCGCACTGCTATCATCTTGTCCACTTTGGATG	1169
OY	241	catgcaataactcatatcttgacagggaaaaatgatbaatccctagttgtctccttgaggagt	300
Db	1170	CATGCAAAATACTCTATTTTGGACAGGAAAAATGATTAACCTAGTTGTCCTGAGAGACTT	1229
OY	301	ggagccactgtctgtgtgacttacttcacccataccagtatgtctcgaatgggggtggaatt	360
Db	1230	GGAGTCACCTGTCTTGTGGACTTCTTACCCCAACGATATGTCTGATGGGGGTGGACTT	1289
OY	361	caaggtcagggagaaagaaaacaagataaaggaaagcaatctcccaactacccggggagat	420
Db	1290	CAGATTCAGGCCAAGAAAACATGTTAAAGAGTATTTCTCCAACTCACCCGGGTACAT	1349
OY	421	agcaccccttagcccttcaaaaggactagtctctcctcaaaactaatgaacccctccgtacc	480
Db	1350	GGCACTCTAGCCCCCTACAAAGAGCTATGCTCTCAAACTACATGAAGAACCTCGGTAC	1409
OY	481	catactgcctgtgtgagctatttaatacacccctcaactcgcgtccatgaaggtctcagc	540
Db	1410	CATCTGCCCGGTGTAACCTATTTAATATCACCCCTCTACGTGGCTCCATGAGGTCTCGACC	1469
OY	541	caaaacccctactaactgtttgatatgctcccccctgcactctgaagcctaacttcaatc	600
Db	1470	CAAAACCTACTAATCTTTGGATATGCTCTCCCTTAACCTTAGGCCATATTTTCAATC	1529
OY	601	ccgtctccctgacaatggaacaactctcagcacagaataaacaacactctcgltttagta	660
Db	1530	CCTGACTCTGAAACATGAGAACCACTTCAGACAGAAATTAACACCACTCTCCGTTTACTA	1589
OY	661	ggacctcttgtttccaatctcgtgaataaacccataactcctaactcaactcgtgtaaatt	720
Db	1590	GGACCTCTGTGTTCAATCTGGAATTAACCCATCTCAACCTCACTCCGTGTAAATTT	1649
OY	721	agcaatatacagacaacaacagctcccaatgatcaggtgggtataacacactcccaacga	780
Db	1650	AGCATATCTACATACACAAACCACTCCCATGATGAGTGGGTAACTCTCCACACAA	1709
OY	781	atagctctgctacccctcaggaataattttgtctgtgtgtaactcagcctatcatgtttg	840
Db	1710	ATAGCTGCTCCTAACCCCTGAGGAATATTTTTGTGTGTGTAACCTCAACCTATGTTGTTTG	1769
OY	841	aatgctctctcagaatactatagtgtctctctctcatctcttagtgccccctatgcaactac	900
Db	1770	AATGGCTCTTTCAGAAATGATGTGCTTCCTCTCAATTTTGTGTGCCCCCTATGACCAATTCAC	1829
OY	901	actgacaagaattatatacatcatgtctgataacgaagcccaacaacaagaatataccatt	960
Db	1830	ACTGACAAGTTTATATACATTTATGTCTATTTCTAAACCCCGCACAAAAGAGTACCCTT	1889
OY	961	cttccctttgttalcagaagagaggtgtcagagactagtagtactgtgcatcttgccagtatc	1020
Db	1890	CTTCCCTTTTGTATTAGGAGCGAGGAGTGCTGTGATGCTGTGATCTGCGCATTTGGCGGTATC	1949
OY	1021	acaaccttactcagttctactatacaactctccaagaataaagtgtgacatggaacag	1080
Db	1950	ACAACCTCTTACTCGTTCTACTCAAACTATTCACAAAGACTTAATGCGGACATGGAAAGG	2009
OY	1081	gtcacctgactccctcgtgcacactctgcaaatcaaatctcaactccctagcaagtaactct	1140
Db	2010	GTCGCGAGACTCCCGGTGCTACCTTGCAAGATCAACACTTACTCCCTAGCACACATATCTCTT	2069
OY	1141	caaaatcgaaagactttagactgtctaaacccgcaaaagaggggaaactgtttatttta	1200
Db	2070	CAAAATCGAAGAGCTTTAGACTTGTCTTAACCGCTGAAGAGAGGGGGAACCTGTTATTTTAA	2129

RESULT	10
AX007999	
LOCUS	2055 bp DNA linear PAT 06-SEP-2000
DEFINITION	Sequence 22 from Patent WO9967395.
ACCESSION	AX007999
VERSION	AX007999.1 GI:995696
KEYWORDS	
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 2055)
TITLE	Perin,J.P., Rieger,F. and Alliel,P.M. Nucleic sequence and deduced protein sequence family with human endogenous retroviral motifs, and their uses Patent: WO 9967395-A 22 29-DEC-1999; INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER FRANCOIS (FR); ALLIEL PATRICK M (FR)
JOURNAL	Location/Qualifiers
FEATURES	1..2055
Source	/organism="Homo sapiens" /db_xref="taxon:9606" <1..87 /note="unnamed protein product" /codon_start=1 /protein_id="CAC07392.1" /db_xref="GI:995697" /translation="PKRNLVADITSLAKYQGVLTIGTYP"
CDS	
BASE COUNT	576 a 574 c 376 g 529 t
ORIGIN	

Query Match	Similarity	90.1%	Score 133.8	DB 6	Length 2055
Best Local	Similarity	93.8%	Pred. No. 0		
Matches 1389	Conservative	0	Mismatches 92	Indels	Gaps
QY	1 atggccctcccttaccatactcttctcttactgtctcttaccaccttgcctact	60			
Db	391 ATGGCCCTCCCTTATCATATTTTCTCTTACTGTCTTATACCCCTTTCACCTCACT	450			
QY	61 gcaacctccatctgtctgttaacaacagtagctcccttaccagaagttctataga	120			
Db	451 GCACCCCTCCATGGCCGCTATGACACATACCTCCCTTACCAAGAGTTTATGAGA	510			
QY	121 acggcgcttccctgggaatatigatgcccatatatagatgttatcgaaggaa	180			
Db	511 ATGCAAGCCTCCGGAAATTTATGATGCCCATGTATAGAGTCTTCTTAAGGAACCCC	570			
QY	181 accttactgtcccaacaccata tgcgccgcaactgtataaacttgcacacttgc	240			
Db	571 ACCTTACTGTCCACACCCATATGCCCCGCAACTGCTATCACTGTGCCACTCTTTG	630			
QY	241 catgcaataactcattatitgacagggaaatgatatacctagtgtccvtgaag	300			
Db	631 CATGCAATATCATATTATTTGACACAGAAAAAGATTAACTAGTTGTCTTGAGAC	690			
QY	301 ggaagcactgtctgttggactcttaccacatacagafatcttgatgggtga	360			
Db	691 GGAAGTACTGTGTGTGGACTTACTTACCCAAATCGTATGTCTGATGGGGTGA	750			
QY	361 caaaggtcgaagcaagaaaaacaagttaaagaaagcaatctcccaactgac	420			
Db	751 CAAGATTCAGGCAAGGAAGAAACATGTAAAGAAGTATTTCCCAATCACC	810			
QY	421 agcacccttagcccttacaagaagactagttcttcaaaaactaataga	480			
Db	811 GGCACCTTAGCCCTTACAAGAGACTAGATCTCTAAACTACATGAACCCCT	870			
QY	481 catctatgcctgttgagccttatattaccaccctactcggctcctatgag	540			
Db	871 CATCTATGCCTGTATAGCCTATTATTATACACCCCTACTAGGCTCATAGGCT	930			

Oy	541	caaaacctactaaactggttgagatgagctcccccgcaccttcaagggcataacttcaatc	600
Db	931	CAAAAACCTACTAACTGTTGGATATGCTTCCCCCTGAAACTTCAGGCCATATGTTTCAATC	990
Oy	601	cctgtccctgacaatatggaacacttcagacagaaataaacaacacttcggtttagta	660
Db	991	CGTGAACCTGGAACAATGGAACAACTTTCAGCAAGAAATAAACACCACTTGGCTTTTAACTA	1050
Oy	661	ggacctcttgttccaatcttggaataaaccataccttaaacctcactctgtgtaaatc	720
Db	1051	GGACCTCTTGTTTCCAAATCTGGAAATTAACCCATACCTCAAAACCTCACTGTTGTAATAATT	1110
Oy	721	agcaatattatagacaacaacagcctcccacatgacatcaggttggttaaacctcccacaga	780
Db	1111	AGCAATFACTACATACACAAACCAACTCCCAATAGCATCAGAGTGGGTACTCTTCCACACAA	1170
Oy	781	atagtctgcctacccctcaggaataattttctgtctgtgtacctcagacctatcatgtttg	840
Db	1171	ATAGTCTGCCTAACCTCGAGGAATATTTTTTGTGTGTGTACTCGCCTATGCGTTGTTTG	1230
Oy	841	aatgctcttcaagaatcattatgtctctctcatcattcttaatgtgcccctatgacatctac	900
Db	1231	AATGCTCTTTAGAAATCATATGCTATGCTTCCTCATCTTATAGTCCCTCATATGACCATATC	1290
Oy	901	actgacaagaatttataaatcatgtgttactatgccccacaanaaagaattaccact	960
Db	1291	ACTGACACAGATTTATACAGTATATCTCATATCTTAAGCCCCGCAACAAAGATACCCAAAT	1350
Oy	961	cttcccttctgtatcacagcagagtgctctagcagactagatgactgcatctggcagatc	1020
Db	1351	CTTCCCTTTTGTATATAGGAGACAGAACTGCTAGGTGCACATAGTACGTGGCATTTGGCGGTATC	1410
Oy	1021	acaacctctactcagttctactacaaactatctcaagaataaatrgtfgacatggaaacag	1080
Db	1411	ACAACTCTTACTCAATCTTACTACTCAAACTATCTCAAGAACTAAAGGGGACATGGAAACGG	1470
Oy	1081	gtcactgtctccctgtgtacactctgcaagaataacacttaactccctctagcagcaatgctct	1140
Db	1471	GTCGCCGACTCCCTGTGTACCTTGCAAGATCAACTTAACTCCCTAGACACACTATGCTCTT	1530
Oy	1141	caaaatcgagaagactcttagactgtctaaacgcacaaagaagggggaacactgttatttlla	1200
Db	1531	CAAAATCGAAGAGCTTTTAGACTTGCTTAACCGCTGAAGAGGGGGAAACCTGTTATTTTTTA	1590
Oy	1201	ggagaagaacgctgttattatggttaatcaatccagaatlttcaactgagaagaatgaaga	1260
Db	1591	GGGGAAGAATGCTGTTATTTATTTATTTAATCAATCCGAATTCGTCACTGAGAAAGTTTAAAGAA	1650
Oy	1261	attggaagtcgaatacaatltgagacagagagagacttcaaaaacacgaacgcgtggggcctc	1320
Db	1651	ATTTCGATGTCGATACAAACGTAGACAGAGGGAGCTTCCAAACACTGAGAACCTCGGGGGCTTC	1710
Oy	1321	ctcagccaatggaatgccttggtgtctccctctcttagagacctctagcagctctaaatattg	1380
Db	1711	CTCAGCCAATGGAATGCCCGATGTTCCCTCTTTAGAGACCTCTAGACGCTATTAATTAATG	1770
Oy	1381	ttactccctcttgagacctgtatctttaacctcctcttgtaagtttgctcttcagaatt	1440
Db	1771	CTATCTCCTCTTTGGACCTGTATCTTTTAACTCTCTTGTATTAATCTTGTCTCTCCAGAAATC	1830
Oy	1441	gaagctgtgaagctcagatggtcttaaaagaatgagaccoca	1481
Db	1831	GAAAGCTGTAAACTACAAATGGAGCCCAAGATGCAAGTCCAA	1871

RESULT	11			
AX007978				
LOCUS	AX007978	2599 bp	DNA	
DEFINITION	Sequence 1 from Patent WO967395.		linear	PAT 06-SEP-2000
ACCESSION	AX007978			
VERSION	AX007978.1	GI:9995675		
KEYWORDS	.			

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 2599)
 AUTHORS Perin,J.P., Rieger,F. and Alliel,P.M.
 TITLE Nucleic sequence and deduced protein sequence family with human
 endogenous retroviral motifs, and their uses
 JOURNAL INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
 FRANCOIS (FR); ALLIEL PATRICK M (FR)
 FEATURES
 source 1..2599
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 744 a 718 c 495 g 642 t
 ORIGIN

Query Match 90.1%; Score 1333.8; DB 6; Length 2599;
 Best Local Similarity 93.8%; Pred. No. 0;
 Matches 1389; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 1 atggccctccctatcatcttcttcttactgttcttccccccttgcctcaact 60
 DB 929 ATGGCCCTCCCTATCATATTTTCTCTTACTGTTCTTTACCCCTTTCACTCTCACT 988
 QY 61 gaccccccctcctatgctctgtatcaaacagtaagccctaccacagagttctatgaaga 120
 DB 989 GCACCCCTCCATGCCGCGTATGACACAGTACCTCCCTTACCAGAGTTTATGAGAGA 1048
 QY 121 acgcggctctctgaaataatgatgcccacatataagagtttactaaggaaactcc 180
 DB 1049 ATGCAGCGCCCGGAATATTTGATGCCCATGCTGATAGAGTCTTTTAAAGGAACCC 1108
 QY 181 acctcactgcccacacacatataagcccgcaactgtataacttgcacacttttgcatag 240
 DB 1109 ACCTTACTGCCCACACCCATATGCCGCCCACTGCTATCACTGCGCCCTTTGCAAG 1168
 QY 241 catgcaaatcatcatatgtgagcaggaagaaatgatatccctagttgtcctgaggaact 300
 DB 1169 CATGCAAAATCATTTATTTGACAGAGAAATGATTAATCCATAGTTGTCTGAGAGACT 1228
 QY 301 ggaagcaactgtctgtggaacttaccacataaccagatgtctgagtgaggatc 360
 DB 1229 GGAGTCACTGCTGTGTGACTTACTTCAACCAACCTGATGTCTGATGGGGTGAGATT 1288
 QY 361 caaggtcgaagcagaagaaacaagtaaggaagcaatctcccaactgacccgggaact 420
 DB 1289 CAAGATCAGGCAAGAAACATGTAAGAGATTAATCTCCCAACTCACCCGGGTACT 1348
 QY 421 agcacccttagccctacaaaggaactagttctctcaaaatcatatgaaccctcgtacc 480
 DB 1349 GGCACCTTAGCCCTACAAAGACTAGATCTCTCAAAATCATAGAAACCTCGGTACC 1408
 QY 481 catactgcctgtgtgagcctatttaataccacctcaactggctccatgaggtctcagcc 540
 DB 1409 CATACTGCGCTGTGTAAGCTATTATTAATCAACCTCACTGGGCTCCATGAGGTCTCG 1468
 QY 541 caaaacccctactaaactgttggatgtgctcccccctgcaactttagggcctaacttcaact 600
 DB 1469 CAAAACCCCTACTACTGTGGATATGCTCCCTCGAAGCTTGAAGCCATATGTTCAATC 1528
 QY 601 cctgttccctgaaatgaaacacttcagcacagaataaacaacacttcgctttagta 660
 DB 1529 CCTGTACCTGAAATGGAACAACCTTCAAGCAAGAAATTAACACACTTCCTTTAGTA 1588
 QY 661 ggaactcttgttccaacttggaaataacccaatccctcaaacctcagctgtgtaaattc 720
 DB 1589 GGACCTCTGTGTTCAATCTGGAATTAACCCATACCTCAAACTCAACCTGATTAATTT 1648
 QY 721 agcaatctatatgacacacacagctcccaatgcatcaggtgtggttaacactcccaaga 780

DB 1649 ACCAATACTACTACACCAACCACTCCCAATGATCAGGTGGGTAACTCTCCACACAA 1708
 QY 781 atagctgcctaccctcagaatatttttgcgtgtggaactcagcactatctgttg 840
 DB 1709 ATAGCTGCCTACCTCAGAGATATTTTGTCTGTGTGACTCAGCCATGCTGTGTTG 1768
 QY 841 aatgctcttcgaatctatgtcttccctcacttctttagtgcacctatgacactac 900
 DB 1769 AATGCTCTTCAAGATCTATGTGCTCCCTCATCTTATGTAGTCCCTATGACATCTAC 1828
 QY 901 actgacaaagattttacatcatatgtcgtactaagcccccaaaaagatgacctt 960
 DB 1829 ACTGAACAGATTTATATGATGATGATATTAAGCCCCGCAACAAAGATATCCATT 1888
 QY 961 ctctcttctgtatcagaacagagtgctaggaactgactgactgtgacatc 1020
 DB 1889 CTTCTTTTGTATATAGAGACAGAGATGCTAGTGCATGATGACATTTGGCGGTATC 1948
 QY 1021 acaactctactagtttctactacaaactatctcaagaataatggtgacatgaaag 1080
 DB 1949 ACAACCTTACTCACTTACTACAAACTATCTCAAGAACTAAATGGGACATGGAACG 2008
 QY 1081 gtcactgactcctgtgtacacttgcagaatcaactaactcctctagcagagtagtcc 1140
 DB 2009 GTGCGGACTCCCTGCTACCTTGCAAGATCACTTAACCTCTAGCAGCATGCTCTT 2068
 QY 1141 caaaatcgaagctcttagactgtctaaaccgcaaaagagggagaaactgttatttta 1200
 DB 2069 CAAATTCGAAGAGCTTTAGACTTGCTTAACCGCTGAAGAGGGGAAACCTGTTATTTTA 2128
 QY 1201 ggaagaagacgtgttattatgttaataccaaatgttcaactgtgagaagttaagaa 1260
 DB 2129 GGGGAAGATGCTGTTATATGTTATATCAATCCGAATCGTCACTGAGAAATTAAGTA 2188
 QY 1261 attcgagatcgaatacaaatgtgagcagagagcttcaaaaacagcagacgctgggccc 1320
 DB 2189 ATTGAGATCGAATACAACTGAGACAGAGAGCTTCGAATCAGTGAGACCTGGGGCTC 2248
 QY 1321 ctcaagcaatgatatccctgtgtctccctctcttagaagcctctagcagctcctaattg 1380
 DB 2249 CTTCAGCAATGATGCCCGGATTCCTCCCTCTTAGAGACCTGACGATTAATATTG 2308
 QY 1381 ttaactcctcttggaacctgtatctttaacctctctgttaagtgtctctcagaatt 1440
 DB 2309 CTATCTCTCTTGGACCCGTATCTTTAAACCTCTTGTAACTTGTCTCTTCAGAAATC 2368
 QY 1441 gaaactgtaaagctacagatgtcttacaatgtgaacccca 1481
 DB 2369 GAAGCTGTAACTCAAAATGAGCCCAAGATGACTCCAA 2409

RESULT 12
 AX007980 10499 bp DNA Linear PAT 06-SEP-2000
 LOCUS AX007980
 DEFINITION Sequence 3 from Patent WO967395.
 ACCESSION AX007980
 VERSION AX007980.1 GI:9995677
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 10499)
 AUTHORS Perin,J.P., Rieger,F. and Alliel,P.M.
 TITLE Nucleic sequence and deduced protein sequence family with human
 endogenous retroviral motifs, and their uses
 JOURNAL INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
 FRANCOIS (FR); ALLIEL PATRICK M (FR)
 FEATURES
 source 1..10499
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

QY 121 acggcgcttcctggaataatgagcccaataatagagttcttaagggaactcc 180
 |||||
 Db 35999 ATGCAGCGCTCCGGAAATATTGATGCCCATGTATAGAGCTTTCTTAAGGAGAACCCCC 36058
 QY 181 acctcactgcccaacacacataatgccccgaactgctataacttcgcaactcttgcatg 240
 |||||
 Db 36059 ACCTTTACATGCCCCACACCCCATATGCCCCGCAACTGTATCACTCTGCCACTCTTTTGCAAG 36118
 QY 241 catgcaaatcaactcaatctatgagcaggaataatgataactcagttgtccctggaagact 300
 |||||
 Db 36119 CATGCAAAATACATCATTTATTTGGACAGAGAAAATGATTAATCCATAGTTGTCTTGAGAGACTT 36178
 QY 301 ggaagcaactgctgttggaacttaacttaaccgaatgctgtagtgggggtggaatt 360
 |||||
 Db 36179 GGAGTACACTGTGTGTGGACTTACTTACACCCAACTGGTATGTGTGATGGGGGTGAGATT 36238
 QY 361 caaagtcaggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 420
 |||||
 Db 36239 CAAGATCAGAGCAAG 36298
 QY 421 agcaccctcagccctcacaagaagactagttctcacaactcaatgaaacctccgtacc 480
 |||||
 Db 36299 GGCACCTCTAGCCCTTACAAAGAGACTAGATCTCTCAAAACTACATGAACCTCCGTACC 36358
 QY 481 caatactcgctggtagagcctatlaataaccacccactcagctcgctcagagtcacagcc 540
 |||||
 Db 36359 CATACTCGCCTGTGAAGCCTATTATTAACACACCTCAGTGGGCTCATAGAGGTCTCGGCC 36418
 QY 541 caaaacccctacaaatgtgtgagttgctgcctccctcagctcagctcagctcagctcagct 600
 |||||
 Db 36419 CAAAACCCCTACTACATGTGTGATGATGCTCCCTCGAAGCTTCAAGCCATATGTTTCAATC 36478
 QY 601 cctgttcctgaaacaatggaacaactcagcagagaataaacaacactcagctcagctcagct 660
 |||||
 Db 36479 CCTGTACCTGAGCAATGAGACAACTTCAAGACAGAAATTAACACACTTCCGTTTAAAGA 36538
 QY 661 ggaactctgtttccaatctggaataacccaatacctcaaacctcagctcagctcagctcagct 720
 |||||
 Db 36539 GGACCTCTGTGTTCCAACTCGGAATAAACCCATACCTCAACCTCAACCTCAACCTCAACCT 36598
 QY 721 agcaatcactatgagac 780
 |||||
 Db 36599 AGCAATACTACTATACACAAAC 36658
 QY 781 atagctcgcctac 840
 |||||
 Db 36659 ATAGTCTGCTACCTCAGAGATATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 36718
 QY 841 aatgacctcagaatcctatgtctcctcctcctcctcctcctcctcctcctcctcctcctcctc 900
 |||||
 Db 36719 AATGCTCTTTCAGAAATCATGTGCTTCCCTCATTTCTTATGCCCCCTATGACCATCTAC 36778
 QY 901 actgagacagatattaataatcagtcagtaagcccaacacacacacacacacacacacacac 960
 |||||
 Db 36779 ACTGAAACAAGATTATATACGTTATCATATCTAAGCCCCCGCAACAAAAGATATCCCAT 36838
 QY 961 ctctccttctgtatcagaagcagagtgctagcagactagagctgagctgagctgagctgagct 1020
 |||||
 Db 36839 CTTCCTTTTGTATATAGACAGAGAGTGTGATGATGATGATGATGATGATGATGATGATGATG 36898
 QY 1021 acaacctactcagctcactac 1080
 |||||
 Db 36899 ACAACCTCTACTCACTTACTACTCAAAACATATCTCAAAACATTAATGGGACATGCAACGG 36958
 QY 1081 gtcactgtcctcctgtgtaacctgtgagaacttaactcctcagcagcagtaactcctc 1140
 |||||
 Db 36959 GTTCGCGAGATCCTCTGTGTACCTTGCAGAGATCACTTAATCTCCCTAGCAGACATATCTCT 37018
 QY 1141 caaaatcgaagaagcttgaactgtctaaaccgcaaaagaggggaacctgtctatttta 1200
 |||||
 Db 37019 CAAATTCGAAGAGCTTTAGACTTGTCTAACCCGTGAAGAGGGGGAACCTGTTTATTTTTA 37078

QY 1201 ggaagaagaacgtgtattatgtaatcaatccagaattgtcactcgtgagaagtaagaag 1260
 |||||
 Db 37079 GGGGAAGAATGCTGTTATTAATGTTAATCAATCCGGAATCGTCACAGAAAGTTAAAGAA 37138
 QY 1261 attgagatcgaatacaatgtgagcagaagaggttcgaacacacagacgcgtggggtc 1320
 |||||
 Db 37139 ATTGAGATCGAATACAAAGTAGAGACAGAGAGCTTCGAAGACATGGAACCTGGGGCCTC 37198
 QY 1321 ctcaagcaatgagtgccctgggtcttccctcctcctcctcctcctcctcctcctcctcctcctc 1380
 |||||
 Db 37199 CTCACCAATGATGATGCTCGATTCCTCCCTCTTAGAGACCTGTAGACATATATTTG 37258
 QY 1381 ttaactcctcttggaacctgtatccttaacctcctctgttaagttgtctcctcagaatt 1440
 |||||
 Db 37259 CTACTCTCTTTTGGACCTGTATCTTTAAACCTCTTGTAACTTGTCTTCCAGAAATC 37318
 QY 1441 gaagctgtgaagctacagatgctcttaacaagaagaccoca 1481
 |||||
 Db 37319 GAAGCTGTAAACTACAAATGAGGCCCAAGATGAGTCCAA 37359

RESULT 14

HSAC000064

LOCUS HSAC000064 56093 bp DNA linear PRI 13-NOV-1996

DEFINITION Human BAC clone RG083M05 from 7q21-7q22, complete sequence.

ACCESSION AC000064

VERSION AC000064.1 GI:1669369

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 56093)

AUTHORS Pauley,A.

TITLE The sequence of H. sapiens BAC clone RG083M05

REFERENCE Unpublished (1996)

AUTHORS 2 (bases 1 to 56093)

TITLE Waterston,R.

JOURNAL Direct Submision

Submitted (13-NOV-1996)

Genome Sequencing Center

Department of Genetics, Washington University

St. Louis, MO 63108, USA

e-mail: saplens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. For references see: Shituya et al., Proc. Natl. Acad. Sci. 89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996). VECTOR: pBELO Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The orientation of this clone is unknown. Actual start of this clone is at base position 1 of H_RG083M05; actual end is at 56093 of H_RG083M05

This clone contains STS SWS1725.

FEATURES Location/Qualifiers

Source 1..56093

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-7q22"
/clone="H_RG083M05"
/cloned_lib="CIRB-978Sk-B"
/complement(838..1131)
/rpt_family="ALU"
<1360..16971
/gene="WUGSC:H_RG083M05.1"
join(<1360..1503,4181..4370,4587..4774,6422..6556,
9483..9547,11631..11773,11864..12021,13131..13296,
14885..14988,16349..16546,16837..16971)
/gene="WUGSC:H_RG083M05.1"
/Note="ATPase, strong similarity to peroxisome
biosynthesis protein Pasi (PID:g1172019); coded for by
human cDNA C04279 (NID:g1467530)"
/codon_start=1
/protein_id="AAB46346.1"
/translation="KRLNIOKTLFVAFSEAVWVQPSVYLDDDLIAGLPAPREHEH
SPDAERCEICLVNKNKIDCDINKFTDLDQVAKETGCFARDDTVLVDRATHSRRL
SHOSISTRKLVLTLDFOKALRGFLPASLRVNLHPRDLGMDTIGLHAEVQIIMSD
T10LPKAVCLKEKYPFLANLPIRQRTGILLYPGTGKTLGAGYIARESRMFTSV
KGBELSKYIGASEQAVRDI FIRAQAKPCILFEDDESIAPRGHDNTGVDRVVNO
LTQLDVGSLGVVYLAATSRPDLIDPALLRGRLDKCYCPPOQVITSYLESKTO
OMLHSLVRLTELAVLSDSLPLADVDIOHVASVYDFTGADLKALVNAOLEALHG
MLSKSEILPDESKRMKRLYRGSSYESELNGTSSDLSQCLSRPSSMTQDLPRVP
KRDQLFSPPVLTASQEGCELTQERDQLRADIISIIKGRYRSQSGEDSMQPPPI
KRLAISQSHLMTALGHTRPISSEDDMKNAEL"
/complement(4948..5130)
/rpt_family="ALU"
complement(6581..7133)
/rpt_family="L1"
complement(7767..8037)
/rpt_family="ALU"
complement(8186..8472)
/rpt_family="ALU"
8473..8625
/gene="WUGSC:H_RG083M05.1"
/Note="match to human 3' EST H75782 (NID:g1049794), bases
287-444"
/gene="WUGSC:H_RG083M05.1"
/Note="match to human 5' EST H75921 (NID:g1050050), bases
21-348"
9481..9547
/gene="WUGSC:H_RG083M05.1"
/Note="match to human 5' EST N22627 (NID:g1130501), bases
276-343"
complement(12612..12907)
/rpt_family="ALU"
13670..13793
/gene="WUGSC:H_RG083M05.1"
/Note="match to human 5' EST H41382 (NID:g17434), bases
143-266"
13794..13877
/rpt_family="ALU"
13878..13906
/gene="WUGSC:H_RG083M05.1"
/Note="match to human 5' EST H41382 (NID:g917434), bases
30-58"
13907..14104
/rpt_family="ALU"
complement(14110..14137)
/rpt_family="L1"
complement(15618..15907)
/rpt_family="ALU"
17327..17522
/rpt_family="ALU"
18667..19235
/Note="match to human fetal brain 5' EST D61494
(NID:g970409), bases 1-255, and to human 3' EST R07476

```

```

(NID:g759399)"
19550..19670
/rpt_family="ALU"
21507..37303
/Note="similarity to various SS-RNA virus polypeptides;
pseudogene; region of matches and close matches to
multiple human ESTs, see R68740 (NID:g842257)"
37316..37489
/Note="Grail prediction, score = 80"
/evidence=not_experimental
complement(38938..39224)
/rpt_family="ALU"
39225..39707
/Note="match to multiple human ESTs, see N30113
(NID:g1148633)"
39800..40085
/rpt_family="ALU"
complement(40247..40538)
/rpt_family="ALU"
complement(40632..40924)
/rpt_family="ALU"
complement(42283..42891)
/rpt_family="ALU"
complement(45474..45613)
/rpt_family="ALU"
complement(45614..45737)
/Note="match to human 3' EST H48898 (NID:g988738), bases
129-333"
complement(46107..47026)
/Note="match to multiple human ESTs, see N81064
(NID:g1243765), H48897 (NID:g988737), and M78831
(NID:g273146)"
complement(47027..47318)
/rpt_family="ALU"
complement(47365..47782)
/Note="match to multiple human ESTs, see W37495
(NID:g1319089)"
47898..48115
/Note="match to human 5' EST H62306 (NID:g1015138), bases
93-368"
complement(48116..48405)
/rpt_family="ALU"
complement(48406..48584)
/Note="match to human 3' EST N29952 (NID:g1148472), bases
290-455, and 5' EST R12730 (NID:g765806)"
complement(48787..49405)
/rpt_family="ALU"
complement(49406..49534)
/Note="match to human 3' EST R65794 (NID:g838432), bases
309-440"
complement(49638..49672)
/rpt_family="ALU"
complement(49674..49890)
/Note="match to human 3' EST N29952 (NID:g1148472) and 5'
EST N29938 (NID:g1148458), sequences are from opposite
ends of the same clone"
complement(49698..51806)
/gene="WUGSC:H_RG083M05.2"
complement(join(49698..49888,51575..51806))
/gene="WUGSC:H_RG083M05.2"
/Note="coded for by human cDNAs M37389 (NID:g1319205),
R65891 (NID:g838529), R65794 (NID:g938432) and R65794
(NID:g838432)"
/codon_start=1
/protein_id="AAB46345.1"
/db_xref="GI:1669370"
/translation="MFYFYQCGIIFCFPGVYVYQIGDVYSVIDEODGKPYAQIRGF
IDQVCEKSAALFWLPIYSSPRDQDPASVYIGPEEDLPKMEYLEFVCHAPSEYFK
SRSSPPTVPTREKGIYTHVGPPTPAITIKESVANHL"
complement(51576..51758)
/gene="WUGSC:H_RG083M05.2"
/Note="Grail prediction, score = 86"
/evidence=not_experimental

```

repeat_region	complement(52052..52329)
misc_feature	<p> /fp_family="LI" 55557..55843 /note="match to human EST M79192 (NID:q273505) base 2-289" </p>

Query Match	90.1%;	Score 1333.8;	DB 9;	Length 56093;
Best Local Similarity	93.8%;	Pred. NO. 0;		
Matches 1389; Conservative	0;	Mismatches 92;	Indels 0;	Gaps 0;

Oy	1	atggccctccacatacaatacttctctctaaatgctctcctaaccctctgcctcaact	60
Db	35879	ATGGCCCTCCCTATCATATTTTCTCTTAACTGTCTTTTAACTCCTCTTACCTCTCACT	353938
Oy	61	gaacccctccatagctgctgatacaaccagtagctcccttaaccaagagtctctatgaaga	120
Db	35939	GCACCCCTCCATGCGCGCTGTGTGACAGAGTAGCTCCCTTACCAAGAGTTCTATGGAGA	359988
Oy	121	acggcgcttctctggaataatgatgccccatataaagagtcttctcaaggaaaccctc	180
Db	35999	ATGCAGCGCTCCGGGAAATATTTGATATCCCCCATCTGTATAGAGCTTTTAAAGGAACCCCC	360588
Oy	181	acccttcaatgcccacaccataatgcccgccaatgctgataactctgcaactcttgaatg	240
Db	36059	ACCTTCACTGCCACACCCATATGCCCCGGCAACTCTATCACTCTGCCACTCTTTGCATG	361188
Oy	241	catgcaaatactcaatatttgaaagaaggaaatgatataaactagtagtcttccctggaagact	300
Db	36119	CATGCAAAATACTCATTTATTTAGACAGGAAAAATGATTAATCCTAGTTGTCCTGAGAGACTT	361788
Oy	301	ggagccactgtctgtgtggaacttaaccataaccaglatgctcgaatgggggtggaat	360
Db	36179	GGAGTCACCTGTCTGTGGAGCTTACTTTCACCCCAACTGGATGTCTGATGGGGGTGAGATT	362388
Oy	361	caaggctcaaggcaagaagaanaagtaaagaagcaatctcccaactgcaaccggggaact	420
Db	36239	CANAGTACAGCAGAGAAAAACATTTAAAGAAAGATATCTCCCAACTCACCCGGGTACAT	362988
Oy	421	agcaacccttagcccttacaagaagactagtctctctaaactcatatgaaaaccctccgtatcc	480
Db	36299	GGCACCTCTAGCCCTCTACAAAGAGACTAGATCTCTCAAAATCATATGAACCTCCGTGACC	363588
Oy	481	catactgcgcgtggtgaagcctatttaatacaaccctcaactcggctcgaatgaggtctcagcc	540
Db	36359	CATACTGCCTGTGTAGGCTTATTTAATACCAACCTCTACCTGCGCTCATGAGGTCTCGGCC	364188
Oy	541	caaaaccccttaactatgtgtggaatgctgcctccccctgcgaacttaaggcctaactttcaatc	600
Db	36419	CAAAACCTCTACTACTGTGTGGATATGCTCCCTCCCTGAACTTGAAGCCATATGTTTCAATC	364788
Oy	601	ccgtctcttgaaacaaatggaaacaacttcagacagaataaacaacacttcggttttagta	660
Db	36479	CCTGTACCTTGAAACAAATGGAAACAACTTTCAGACAGAAATTAACACCACTTCGTTTAGTA	365388
Oy	661	ggacctcttgtttccaatctgggaatlaaaccaatacccaaaccttaactgtgttaaatctt	720
Db	36539	GGACCTCTTGTTTCCAAATCTGGAAATTAACCCATACCTCAAAACCTCACTGCTGTAAATTT	365988
Oy	721	agcaatactatagacacaacaacagcgtcccaatgcatcagaatgggtgaataacgtccacaga	780
Db	36599	AGCAATACTACTATACACAAACCAACTGCCATGATCAGGTGGGTAACTCTTCCACACAA	366588
Oy	781	atagctctgactaccctcaggaatatcttctgtctgtgtgtacccagcctatcatgtttg	840
Db	36659	ATAGCTCTGCTTACCTCAAGAAATATTTTTTGTCTGTGGTACTCAGCGTATGCTTGTGG	367188
Oy	841	aatgctcttcagaaatctatgtgtctcctctcatctctagtgccccctatgacaaactac	900
Db	36719	AATGCTCTTTCAGAAATCATATGCTTCCCTCCTCATTTAGTCCCCCTATGACCATCTAC	367788
Oy	901	actgaaacaaagatttatatacatatgtgtgtactaaagcccccaaaaaaagaattaccactt	960
Db	36779	ACTGAAACAAGATTTTATACAGTTATGTCTATATCTAATTAAGCCCCCGCAAAAAAGATATCCCATTT	368388

QY	961	ctccctttgtatcaaggagggagtgctgaagcagactaggctactgaggtcgaatc	1020
Db	36839	cttccctttgtatgagagcagagatgctagtgccactgagctactgagctggcgtatgc	36898
QY	1021	acaacctactcaactcttactacaactatctcaagaataaataatgctgacatgaaacag	1080
Db	36899	acaacctctactcaactttctactacaactatctcaagaactaaatgagacatgaaacag	36958
QY	1081	gtcactgactccctggtcaactctgcaagatcaacttaaccctccatgacgaagtccct	1140
Db	36959	gtcgcgcgactcccttgcgtacacttgcacatgacatcaacttaactccctgacgacgtatgctctt	37018
QY	1141	caaaatcgaaagagctttagactctgtcaacgccaagaagggggaacctgttatcttta	1200
Db	37019	caaaatcgaaagagctttagactctgttaacccgtgaagaagggggaacctgtttatttttta	37078
QY	1201	ggaagaagaagcgtgttatatgttaatcaatccagaactgtgcaactgagaagaattaaaga	1260
Db	37079	ggggaagaagatgctgttatattattattattatcaatccgaatcgtgacactgaagaattaaaga	37138
QY	1261	atctgagatcgaaatacaatgtagagcagaagagcttcaaaaacccgaacgctgaggctc	1320
Db	37139	atttcagatcgcaatgcacagtagagcagaagagcttcaaaaacactgacacctggggccctc	37198
QY	1321	ctcgaacgaatgagtccctggtctccctcttcttagaacctcttagagcgtctaatatgt	1380
Db	37199	cttcacccaatgagatgctccctggtctccctcttcttagaacctcttagagcgtctaatatgtg	37258
QY	1381	ttaacctctcttgagacctgtatcttaacctcctctgttaagtgtctctccagaatt	1440
Db	37259	cttacctctcttggacacctgtatctttaaacctctgttaagtgtctctccagaattc	37318
QY	1441	gaagctgtaaaagctacagatggtcttacaatggaaccca	1481
Db	37319	gaagctgtaaaagctacaaatggaagcccaaaatgacagctccaa	37359
RESULT	15		
LOCUS	AC007566/C	149194 bp	DNA linear PRI 06-FEB-2002
DEFINITION	Homo sapiens chromosome 7 clone CTR-1005, complete sequence.		
ACCESSION	AC007566		
VERSION	AC007566.2	GI:11181861	
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE			
AUTHORS	Waterston,R.H.		
JOURNAL	Unpublished		
REFERENCE			
AUTHORS	Waterston,R.H.		
JOURNAL	Submitted (15-MAY-1999) Genome Sequencing Center, Washington		
REFERENCE			
AUTHORS	University School of Medicine, 4444 Forest Park Parkway, St. Louis,		
JOURNAL	MO 63108, USA		
REFERENCE			
AUTHORS	Waterston,R.		
JOURNAL	Submitted (02-OCT-2000) Department of Genetics, Washington		
REFERENCE			
AUTHORS	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA		
JOURNAL	4 (bases 1 to 149194)		
REFERENCE			
AUTHORS	Waterston,R.H.		
JOURNAL	Submitted (15-NOV-2000) Genome Sequencing Center, Washington		
REFERENCE			
AUTHORS	University School of Medicine, 4444 Forest Park Parkway, St. Louis,		
JOURNAL	MO 63108, USA		
REFERENCE			
AUTHORS	Waterston,R.H.		
JOURNAL	Submitted (03-JAN-2002) Genome Sequencing Center, Washington		

us-09-319-156a-9.rge

Db	853/5	CGTGA	CTGGA	CAATG	GACCA	CTTCAG	CAAGAA	TAACCC	CACTCC	CTTTAG	TA	854/6
Db	853/5	CGTGA	CTGGA	CAATG	GACCA	CTTCAG	CAAGAA	TAACCC	CACTCC	CTTTAG	TA	854/6
Qy	661	ggacctctgtttcc	aatc	ggaat	ccca	tacct	aaac	cttacc	ctgt	gtaaa	ttt	720
Db	854/5	GGACCTCTGTGTT	CCAAAT	CTGGA	AAATTA	CCCTAC	CTCAAA	ACCTC	ACCTG	CTGTA	AAATTT	854/6

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

Oy	781	atagctgcctacccctcagaataattttgtcgtgtgtaacctagaccatcatgtttg	840
Db	85345	ATAGCTGCTCCTACCTCAGAAATATTTTGTGCTGTGTAAGCTCAGCCATGCTGTTC	85286
Oy	841	aatgagcttcagaactatgtgtcctcctcatcttagtgcgccctatgacatctac	900

901 actgaacaagatttatatacatcatgtgtcgtaacctagcccccaacaanaagagtaaccatt 960

Db 85165 CTTCCCTTTTGTATAGCAGCAGAGTCTCAGTGCACCTAGCTACTGCGATTGGCGGTTTC 85106

Db 85105 ACAACCTCTACTCAGTTCTACTACAAACTATCTCAGAACTAANTGGGGACATGGAACGG 85046

Db 85045 GTCCGGACCTCCCTGGTCACCTTGCACATCACTTAATCTCCCTAGCAGCAGTAGTCCTT 849886

QY 1141 caaaatcgaaagccttagactgtcaaccgccaaaagaggggaaacctgtatttllta 1200
|||||

QY 1201 ggagaaagacgctgtattatgttaatcaatccagaatttcactgagaaagttaaagaa 1260

Db 84925 GGGGAGAAGCTGTTATATGCTTAATCATATCCGGAATCGTCACGTAGAAAGTTAAAGAA 84866

Db 84865 ATTCGAGTCGATACACGTCGAGCAGAGGAGCTTCGAAACACTGGACCCCTGGGGCCTC 84806

Qy 1321 ctccagccaatgatgcctcgtgtctccctctctagacctctagacgtctcaattg 1380

QY 1381 ttactcctctttgaccctglatcctttaaocctccttqtltaagtttgcctctccagaatt 1440

Db 84745 CTACTCCTCTTTGGACCGTATCTTTAACTCCTCTTGTTAACTTGTCTCTTCCAGAAATC 84686

Db 84685 GAACGTGTAAACTTACAATGTGAGCCCCAGATGCACTCCAA 84645

Job time: 11276 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 08:18:31 ; Search time 3276.58 Seconds
(Without alignments)
5474.444 Million cell updates/sec

Title: US-09-319-156a-12
Perfect score: 1329
Sequence: 1 tcaaatcgaagacttag.....gcaagatccccagtaaca 1329

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estdb:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vtc:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	593.4	44.7	771	10	BI087886 602852690
2	480.8	36.2	679	12	AG113694 Pan trogl
3	475.2	35.8	683	12	AG134524 Pan trogl
4	474	35.7	669	12	AG049953 Pan trogl
5	473	35.6	681	12	AG116971 Pan trogl
6	469.6	35.3	710	12	AG096341 Pan trogl
7	456.2	34.3	670	12	AG072852 Pan trogl
8	455	34.2	712	12	AG099717 Pan trogl
9	439.2	33.0	689	12	AG121669 Pan trogl
10	438.8	33.0	921	12	BH149565 EMTG48TR
11	435.6	32.8	679	12	AG076758 Pan trogl
12	433.6	32.6	683	12	AG102951 Pan trogl
13	430.6	32.4	611	12	AG038171 Pan trogl
14	429.4	32.3	736	12	AG086046 Pan trogl
15	427.6	32.2	714	12	AG0238712 RPII1-64
16	424.4	31.9	697	12	AG108186 Pan trogl
17	423.8	31.9	675	12	AG117095 Pan trogl

RESULT	1	771 bp	MRNA	linear	EST 20-JUN-2001
BI087886	602852690F1 NIH_MGC_10	Homo sapiens	CDNA	clone	IMAGE:4993894 5'
LOCUS	BI087886				
DEFINITION	602852690F1 NIH_MGC_10	Homo sapiens	CDNA	clone	IMAGE:4993894 5'
ACCESSION	BI087886				
VERSION	BI087886.1	GI:14506216			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	NIH-MGC http://mgc.ncl.nih.gov/				
AUTHORS	1 (bases 1 to 771)				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue-Procurement: ATCC CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: Incyte Genomics, Inc. DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: http://image.llnl.gov Plate: LAM11015 row: d column: 23 High quality sequence stop: 762. Location/Qualifiers 1..771 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4993894" /clone_id="NIH_MGC_10" /cell_line="MGC36" /lab_host="DH10B" /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.5 kb. Library prepared by Life Technologies."				

ALIGNMENTS

BASE COUNT

225 a 206 c 166 g 174 t

ORIGIN

Query Match 44.7% Score 593.4: DB 10; Length 771;
 Best Local Similarity 89.2% Pred. No. 2e-159;
 Matches 688; Conservative 0; Mismatches 68; Indels 15; Gaps 4;

QY 107 gagaagaatgaagaatctgagatcgaatataatgtagagagagagacccctcaaacact 166
 DB 1 GAGAAAGTTAAGAAATTCGATCGAATATACAGTAGAGAGAGAGCTTCGAAACACT 60
 QY 167 gacacctggggcctctctcaagcgaatgagatgctctccctctcttaagacctca 226
 DB 61 GGACCCCTGGGGCCTCTCTCAAGCAATGAGATGCTTCGATTCCTCCCTCTTACGACTCA 120
 QY 227 gagaatataatcttct 286
 DB 121 GAGACTATATATTTGCTACTCTCTTTTGAGACCTGATCTTAACTCTCTTTTAACTTT 180
 QY 287 gtcctctccagaatctgaagctgtlaagctacaatagttcttcaaatggaaccagatg 346
 DB 181 GTCTCTTCAGAAATCGAAGCTGTAATAACTA-----CAATGAGGCCAAGATG 228
 QY 347 cagtcacagcacaataatctaacctgtgacccctggagccgctctgctagactatgctgat 406
 DB 229 CAGTCCAAACTAAGATTCATCCGACAGACCCCTGAGACCCGCTGCTAAGCCAGATCTGAT 288
 QY 407 gtaatgacattgaagtcacccctcccgaggaatctcaactgcaaacccctactacac 466
 DB 289 GTTATGACATCAAAAGCCACCCCTCTGAGAAATCTCAGCTGACACACTTACTACGC 348
 QY 467 tccaattcagtagaagcagtagagcagctgttcagccaacctccccaagctactggg 526
 DB 349 CCCAATTTCAGCAGGAAGCTTAGAGCGGTGCTC--GGCAACCTCCCAAAACACACTTAG 407
 QY 527 tttctctgt-agaaggggtgagctgagagacagagctgctgatttcttaagctgagacta 585
 DB 408 TTTCTCTGTGATGAGTGGGGGAGTGAAGACAGACTACTGATTTCTTACGGCTGACTCA 467
 QY 586 agaattcccaagaacctanctggaggaaggtgacccatccatcttcaaacatgagctgtgcaa 645
 DB 468 ACAATTCCTTAACCTAGCTAGCTGAGAGGTGACACATCCACCTTTAAACAGGGGCTTGCA 527
 QY 646 cttaagctcaaccgcccaatcagaagagctcaataaagttaactcaagcaaaaagga 705
 DB 528 CTAGCTCACACCTGACCAATCAGAGAGCTCACTAAATGCTAATTAGCAAGACAGA 587
 QY 706 gcta-aagaatcagaatcactctatctgtctgagagcagagggaggaagaagattgg 764
 DB 588 GGTACAGAAATAGCCAAATCATCTATGCTGAGAGCACAGCAGGAGGACAAATGATCG 647
 QY 765 gataataaactcagagcattcaagcagcaaacacccctctgggtccctccatctg 824
 DB 648 GATATTAAACCAAGCTTTCGAGCCGCAACGGCAACCCCTTTGGGTCCCTCTTTGT 707
 QY 825 atggagagctctgttctactactatctactactataatcaatcagcaatgca 875
 DB 708 ATGGAGCTCTGTATCATGATGATTTCACTTAAATCTTGAACAATGCA 758

RESULT 2
 AG113694/c AG113694 679 bp DNA linear GSS 03-NOV-2001
 LOCUS Pan troglodytes DNA, clone: PTB-120G11.R, genomic survey sequence.
 DEFINITION AG113694
 ACCESSION AG113694
 VERSION AG113694.1 GI:16734213
 KEYWORDS GSS: GSS (genome survey sequence).
 SOURCE Pan troglodytes male lymphoblast DNA, clone_11b:PTB Chimpanzee Male
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
 REFERENCE 1 (sites)

AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
 Tokok, Y., Watanabe, H. and Sakaki, Y.
 TITLE BAC end sequences of library PTB
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 679)
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
 Tokok, Y., Watanabe, H. and Sakaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Suhiro-chou, Tsukuba, Ibaraki, Japan, 305-8565, Japan
 (E-mail:chimps@sc.riken.go.jp, URL:http://nmp.gsc.riken.go.jp/,
 Tel:81-45-503-9111, Fax:81-45-503-9170)
 Clones are derived from the chimpanzee BAC library PTB This BAC end
 was generated during the Rad process and may have higher chance of
 clone tracking errors.
 PRIMERS
 Sequencing: M13rev
 LIBRARY
 Vector : pKS145
 R.Site 1 : SacI
 R.Site 2 : SacI.
 Location/Qualifiers
 1. 679
 /organism="Pan troglodytes"
 /db_xref="taxon:9598"
 /clone="PTB-120G11.R"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_11b="PTB Chimpanzee Male BAC library"
 BASE COUNT 168 a 177 c 189 g 144 t 1 others

ORIGIN

Query Match 36.2% Score 480.8: DB 12; Length 679;
 Best Local Similarity 86.7% Pred. No. 4.5e-127;
 Matches 552; Conservative 0; Mismatches 85; Indels 2; Gaps 2;

QY 673 gctcaataaactgctcaatcagaagcaaaacagaggtlaagcgaatagccaatcatctatg 732
 DB 678 GCTCGTAAAGATGTAATTAAGCAAAACAGAGGTAAGAAATAGCCATCATCTATTG 619
 QY 733 cctgaagcagacagcggaagcaagagatgtagataataactcagagcattcaagcagca 792
 DB 618 CGTGAGAGCACAGTGTAAAGAAAGGTCGGGATTAAGCCCGGCAATTCAGCGCGCA 559
 QY 793 ac-agcaaacaccttgggttccctcccaatgtatggagagctt-gtttcaactatct 850
 DB 558 ACGGCGAGCCCTTTGGGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 499
 QY 851 cactctatataatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 910
 DB 498 CACTATATTAGATCTTTCAGCTGACACTCTTTCGCGGCTTGTACAGCTCAAAACAGA 439
 QY 911 gctttgttcgcacatccacacactgctgtttgcacacgctcaagaccgctgctgctcc 970
 DB 438 GCTTTGTCTTGGCTTCGCGGCACTGCTGTTTACCGCGCTCCGACCGGCACTGACTTC 379
 QY 971 atcccttggatcagaagagtgctcaactgtctcctgataccagcaggttaccatgccc 1030
 DB 378 ATCCCTCTGGATCCGAGGAGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 319
 QY 1031 actcccgatcagagctaaagagcttgcacatgttctcgtcagatgctgaagctgctgtgtc 1090
 DB 318 GCTCCCGATCGGGCTAAAGGGTGTGCATGTTCTGCAATGCGCTAAGTGCTGCTGCTGCT 259
 QY 1091 ctaatagaactgaacactggttcaactggtgttcaatgttctcttcaatgaacagagcttc 1150
 DB 258 GTAATGAGACTTAACACTAGTGTGAGTTCACAGTCTCTCTCCGTGACCCACGCTTG 199
 QY 1151 taatagagctataaacactcagcagctgagcgaagatccatcctctgtgattctgagagc 1210
 DB 198 TAATGAGACTGTACGCTACCGCATGGCCCAAGATTCCGTTCTTGGCAATCGTGAGGC 139

QY	1271	tttgcgtacgagccaccacatcttggagagctgtg	1307
Db	138	CAAGAACCCCAAGGTCAGGAACACAGAGCTTGGCCACCATCTTGGAAAGTGGCCCGCTGCCA	79
QY	1271	tttgcgtacgagccaccacatcttggagagctgtg <td>1307</td>	1307
Db	78	TTTTGGAAGTGGCCCGCCACCATCTTGGAGACTCTCGC	42
RESULT	3		
LOCUS	AG134524	683 bp	DNA
DEFINITION	Pan troglodytes DNA, clone: PTB-147110.R, genomic survey sequence.		
ACCESSION	AG134524		
VERSION	AG134524.1	GI:16664202	
KEYWORDS	GSS; GSS (genome survey sequence).		
SOURCE	Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male BAC Library Clone:PTB-147110.R.		
ORGANISM	Pan troglodytes		
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Pan.		
AUTHORS	1 (sites)		
TITLE	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,		
JOURNAL	Totoki, Y., Watanabe, H. and Sakaki, Y.		
REFERENCE	BAC end sequences of library PTB		
AUTHORS	Unpublished		
TITLE	2 (bases 1 to 683)		
JOURNAL	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,		
REFERENCE	Totoki, Y., Watanabe, H. and Sakaki, Y.		
AUTHORS	Direct Submission		
TITLE	Submitted (02-AUG-2001) Aaso Fujiyama, The Institute of Physical		
JOURNAL	and Chemical Research (RIKEN), Genomic Sciences Center (GSC);		
REFERENCE	1-7-22 Suehito-Chou, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan		
AUTHORS	(E-mail: chimbes@sc.riken.go.jp, URL: http://bgp.gsc.riken.go.jp/;		
TITLE	Tel:81-45-503-9111, Fax:81-45-503-9170)		
JOURNAL	Clones are derived from the chimpanzee BAC library PTB This BAC end		
REFERENCE	was generated during the R&D process and may have higher chance of		
AUTHORS	clone tracking errors.		
TITLE	PRIMERS		
JOURNAL	Sequencing: M13Rev		
REFERENCE	LIBRARY		
AUTHORS	Vector : pKS145		
TITLE	R.Site 1 : SacI		
JOURNAL	R.Site 2 : SacI.		
REFERENCE	Location/Qualifiers		
AUTHORS	1..683		
TITLE	/organism="Pan troglodytes"		
JOURNAL	/db_xref="taxon:9598"		
REFERENCE	/clone="PTB-147110.R"		
AUTHORS	/sex="male"		
TITLE	/cell_type="lymphoblast"		
JOURNAL	/clone_lib="PTB Chimpanzee Male BAC Library"		
REFERENCE	160 a 201 c 156 g 164 t 2 others		
AUTHORS	ORIGIN		
TITLE	BASE COUNT		
JOURNAL	160 a 201 c 156 g 164 t 2 others		
REFERENCE	Query Match 35.8%; Score 475.2; DB 12; Length 683;		
AUTHORS	Best Local Similarity 89.6%; Pred. No. 1.9e-125;		
TITLE	Matches 554; Conservative 0; Mismatches 60; Indels 4; Gaps 4;		
JOURNAL	QY 666 tcaagagactcctaataatgctcaatcaggcaaaaacagaggttaagcaatagccaatca		725
REFERENCE	Db 58 tgcgnagactctcctaaatgcttaattagcgaacaaacagaggttaagcaatagccaatca		117
AUTHORS	QY 726 tctattgctcgaagacagcgggaagacaaagattggagataataactcaggactca		785
TITLE	Db 118 tctattgctcgaagacagcgggaagacaaagattggagataataactcaggactca		177
JOURNAL	QY 786 gccacacagaaaccccttgggtccctccatgatatggagagctgtttcattc		845
REFERENCE	Db 178 gccacacagaaaccccttgggtccctccatgatatggagagctgtttcattc		236

OY	846	tatttcactcattataatcaatgcgaacgtgacctcttctgtccgttgcttttatltgatgctaa	905
OY	906	gctgagcttttgttgcgcaccacacgtctg- tttgcacacgtcacagaccgcgtctg	964
Db	237	TATTTCACTCATTAATAATCTGGCAACTCACGCTTCTGTGGCCGTGTTTGTAAGCGCTGGA	296
OY	297	GCTAGACATTTCCTCCTGCACATCCACACACGTCGTGTTTTGCCGGCCGTCGACAGACC	356
OY	965	actcaccctcttggatcccagcagagtgtccactgtgtccctcgatcccaagcaggtaacc	1024
Db	357	ACTTCCATCCCTCCAGATCCAGCAGAGGCTCCGCTGCTCTCTGATCCAGCAGAGCGCCC	416
OY	1025	attgccactccccgatccaagctcaaagcgttgcattgtcttcctgcgaagtgaagtgtctgg	1084
Db	417	ATTGGCGCTCTCGAATGGGGCTAAAGGCTTGGCATTTGCTTCGACAGGCTAAAGTGGCTTGG	476
OY	1085	ttgtgccctaatagaactgaacacgtgtcatctgggtgttcattgtctcttcctcatagaaccac	1144
Db	477	TTTGGTCCTATCACTAACCTGAACACTAGTCACTGCTGGTTCCACGGTTCTCTTCCATGAACAC	536
OY	1145	gacctctaataagactataaacactccacgtgcacgtgccccagaattccatcct-tgtatct	1203
Db	537	GACCTTCTATAGAGCTATAACACTCAGTCAGTAGGCGCTAAAGATTCCATTCTTGGGGAATCC	596
OY	1204	gtgagggccaa-gaaccccaaggtcagaagaangtgaagcttgcacacatttgggaagtggcc	1262
Db	597	ATGAGGCCAATGTAACCCACAGTGCAGAAAAACAAGAAGCTTGGCCACCACTTGAAATGGCC	656
OY	1263	cactgcacatttgytagc 1280	
Db	657	CACCACTCTGTGGGAC 674	
RESULT	4		
LOCUS	AG049953	669 bp DNA linear GSS 02-NOV-2001	
DEFINITION	Pan troglodytes DNA, clone: PTB-030D19.F, genomic survey sequence.		
ACCESSION	AG049953		
VERSION	AG049953.1 GI:16586845		
KEYWORDS	GSS; GSS (genome survey sequence); Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male BAC library clone:PTB-030D19.F.		
SOURCE	Pan troglodytes		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.		
REFERENCE	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Tsuchi,Y., Matanabe,H. and Sakaki,Y.		
AUTHORS	BAC end sequences of Library PTB		
TITLE	Unpublished		
JOURNAL	2 (bases 1 to 669)		
REFERENCE	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Tsuchi,Y., Matanabe,H. and Sakaki,Y.		
AUTHORS	Direct Submission		
TITLE	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suhei-ro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimbessgsc.riken.go.jp, URL:http://ngp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)		
JOURNAL	Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the Red process and may have higher chance of clone tracking errors.		
COMMENT	PRIMERS		
	Sequencing: -21M13		
	LIBRARY		
	Vector : pKS145		
	R.Site 1 : SacI		
	R.Site 2 : SacI.		
	Location/Qualifiers		
	1..669		
	/organism="Pan troglodytes"		
	/db_xref="taxon:9598"		
	/clone="PTB-030D19.F"		
FEATURES	source		

BASE COUNT	151 a	185 c	160 g	173 t
ORIGIN	/sex="male" /cell_type="lymphoblast" /clone_lib="PTB Chimpanzee Male BAC Library"			
Query Match	35.7%	Score 474:	DB 12:	Length 669:
Best Local Similarity	86.8%	Pred. No. 4.1e-125:		
Matches 548:	Conservative	0:	Mismatches 71:	Indels 12: Gaps 2:
Qy	669	gagaactcactaaatgctcaatcagcaaaaacaggaagtaaaagcattgcccattcatct	728	
Db	49	GAGACCTCACTAAATGCTGATTAGGCMAAACAGAGAGGTAAAGAATAGCCAAATCATCTT	108	
Qy	729	attgcctgagagacagcggggaaggaacagatctggatataaacctcaggtcattcaagcc	788	
Db	109	ATTGCTGAGAGACAGCAGCAGGAGGAAATGATCGGATATTAACCCAGCATTTGCAGCT	168	
Qy	789	agcaacagcaaacccctcttgggtccctcccatctgtatagggagctctgtttccattctat	848	
Db	169	GGAACAGGCTTACCTTTTGGGTCCCTCCCTCTTTGTATGAGAGCTGTG-----T	218	
Qy	849	ttcactctatbaatcattgcaactgcaactc--ttctgtgctcgtgtttttatagctcaag	906	
Db	219	TTCACTCTATTAAATCTTGCAACTGCACCTCTTTTATGTTCCATGTTGTACAGCTCGAG	278	
Qy	907	ctgaagctttgtctgcacatcccaactgctgttttgcacacgctcaagaacccgctgtcag	966	
Db	279	GTGACCTTTTGTGATTCGGCCGACACAGTGCTGTTTCTGCTGCGACACCCAGCCGCTGAC	338	
Qy	967	ttccatcccttgatgccagcagagtgctcaactgtctctgtatccagcagagtaacct	1026	
Db	339	TTCCATCCCTCTGATCTGGCATCTGGCGGTGTCCTGCTCTCTCATCAGCAGAGTGCCTCAT	398	
Qy	1027	tgccactcccgatcagcgttaaaagctttgcattgtcttcctgcacatgctcaagtacctggtt	1086	
Db	399	TGCTCTCTCTGATTCGGGATAAAGGCTTGCCATTGTTCTGCAAGGCTTAAGTCTCTGGGTT	458	
Qy	1087	tgctcctaataagaaactgcaacaactgctgcaactggttttccattgtctcttccatgaccaca	1146	
Db	459	CGTCTTAATCGAGCTGGAACACTGACATGACATGCGATTCATGTTCTCTTCGGTACATCATG	518	
Qy	1147	cttctcaataagagctataaactcctcaeccgcatggtgcccaagaatctcattcctgttatctgt	1206	
Db	519	CTTCTAAATAGACTTAAACACTACACCGCATGCGCCCAAGATTCATTCCTTGGAAATCCGG	578	
Qy	1207	aggccaacaaaccccgatcagagaagaatggaagcttgcacccatttgggaagtgagccact	1266	
Db	579	AGGCCAAGAACCCCGATCAGAGAACACAGAGGCTTGCCACCAATCTTGGAGATGACACACC	638	
Qy	1267	gccaatttgtagcggccacacacacattctg 1297		
Db	639	ACCAATTTTGAAGACAGCCGCCACCATGATG 669		
RESULT 5				
LOCUS	AG116971	681 bp	DNA	linear
DEFINITION	Pan troglodytes DNA, clone: PTB-124K18.F, genomic survey sequence.			
ACCESSION	AG116971			
VERSION	AG116971.1	GI:16737490		
KEYWORDS	GSS: GSS (genome survey sequence).			
SOURCE	Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male			
ORGANISM	BAC library clone:PTB-124K18.F.			
REFERENCE	Pan troglodytes			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
TITLE	Mammalia; Mammalia; Primates; Catarrhini; Hominiidae; Pan.			
JOURNAL	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,			
	Toko, Y., Watanabe, H. and Sakaki, Y.			
	BAC end sequences of library PTB			
	Unpublished			

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
2 (bases 1 to 681)	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Toki, Y., Watanabe, H. and Sakaki, Y.	Direct Submission	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimpanzee@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170)	Clones are derived from the chimpanzee BAC library PTB. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS	Sequencing: -21M13			
LIBRARY	Vector 1 : pKS145			
	R.Site 1 : SacI			
	R.Site 2 : SacI			
FEATURES	Location/Qualifiers			
source	1..681			
	/organism="Pan troglodytes"			
	/db_xref="taxon:9598"			
	/clone="PTB-124K18.F"			
	/sex="male"			
	/cell_type="lymphoblast"			
	/clone_id="PTB Chimpanzee Male BAC Library"			
BASE COUNT	155 a 207 c 154 g 165 t			
ORIGIN				
Query Match	35.6%	Score 473;	DB 12;	Length 681;
Best Local Similarity	86.8%	Pred. No. 8e-125;		
Matches 538;	Conservative	0;	Mismatches 56;	Indels 12; Gaps 2;
QY	671	gagctcactaaatgycatcaggcaaaaaggaggtaaagcaatgacatcatcat	730	
DB	86	GAGCTCACTAAAAGCTATTATGAGCAAAAACAGAGGTAAGAAATAGCCATATCTAT	145	
QY	731	tgactgagagacacagcgggaagaacagattggatataaactcaagcattcaagccag	790	
DB	146	TGCGTAGAGACGACAGACAGAGGAGCAATGATGGGATATAAACCCAGCATTCAGCCAG	205	
QY	791	caaacagcaacccctttgggtgccctcccatgtatgaggagctctgtttcaacttat	850	
DB	206	CAACGGCTACCCCTTGGGTCCTCCCTCCCTTGTATGGAGCTCTGT-----TTT	255	
QY	851	cacctatataatcatgcaactgtca--ctctctggtccgtgttttatagtcacaagct	908	
DB	256	CACCTATTAAATCTTGTGCACTGCACACTCTCTCTGTCGTCGTTTATAGCTCAAGCT	315	
QY	909	gagctttgttcgscatccacacactgctgtttgtgcacgctcacagacgctcgtcga	968	
DB	316	GAGCTTGGCTTGGCATTCACACACTGCTGTTTGGCGCGCGCGAGACCCGCTGTGACTT	375	
QY	969	ccatccttggatccagcagagtgctcaactgctgctctgataccagcgaggtaccat	1028	
DB	376	CCATCCCTCCAGACTGTGGCAGAGGTCCTCAGTCTCTGTATCCAGAGGCGCCCATTTG	435	
QY	1029	ccactcccgatcagcgttaaaagcttgcatatgtctccgcgaagtgtaagtgccgtgttg	1088	
DB	436	CCGCTCCCGATAGGGCTTAAGGGCTTGGCCATTGTTCTGCAATGAGCTAAGCTGCCAGGTTCA	495	
QY	1089	tcctaatagaacttaaacactggtcaactggtgttcatagttctcttcatagaaccagct	1148	
DB	496	TCTTAATCGACTGTAACACTAGTATACCGGCTTCCATGTTCTTCCGTATACCCAGCGCT	555	
QY	1149	tctaatagaactataaacactcaaccgcatggtgccaaagattccatctccttgglatctgag	1208	
DB	556	TCTTAATAGAGCTATTAAACACTCACCGCATGGGCCAAAGATTCCATTCTTTGAATTCGTGAG	615	
QY	1209	gcacaagaaccccaaggtcagaagaangtgaagctgtgcacacatttgggaagtggcccactgc	1268	
DB	616	GCCTAAGAACCCAGTACAGAGAACACACAAAGGCTTCCACACTCTTTGGAAAGTGGCCCAACAC	675	

OY	911	gctttgttcgcacgaacactgctgttgcacagccgaaccgcgtcgtatccc	970
Db	305	GCTTTTGCTCACAACATCCACCACACTGCTGTATTGCCCGCGTTCAGAACCAATCGCTGACTTCC	364
OY	971	atcccccttgatcccgcaagaagtgtccaatgtgtctctcgtatccaagcaggtaaccattgcc	1030
Db	365	ATMCCCTCCAGATMTGGCAGGGGTGCACACTGTGTCCTCATCCAGGCAAGCACCACTTGGC	424
OY	1031	actccgcgatcaggtctaaagcttgtccattgtctctcgtcatgbgtctaagtcacctggattgc	1090
Db	425	GCTCCCGATCCGGCGCTAAAGGGCTTGCCATGTGTCCTGCACAGCGCTAAGTCCCTGGGTCCGTC	484
OY	1091	cttaatagacgtgaacacgcgfcacgcggtgttcatagtgtctctcttcattaccacaagcttc	1150
Db	485	CTTAATCAAGCTGAACACTAGTACACTGGGGTTCATGGTTCTCTTCATATACCATGAGCTTC	544
OY	1151	tatatagaactataaacctcacccacgcacatgtgcccagaattccattccattcctgtgatcgtgagc	1210
Db	545	TAAATAGACCTTAATACACTCACCAGCATGGCCCCAAGAATTCCATTCTTGGAAATCCGTGAAGC	604
OY	1211	caagaacccccagtcacgaagaangttagagcttgcacacattgtggaagtgccccatgcc	1270
Db	605	CAAGAACCCTCAGCGTCAGAGAACACAGACT-----GCCACCA	641
OY	1271	tttgtgtagcggccacccacccatcttggtagctgtgtgggaagcaagatccccagctaaca	1329
Db	642	TCTTGGAGGACAGCCCTGCACCATCTTGGAACTCTATATAGACAAAGAACCCCGCAGTAAACA	700
RESULT	7		
LOCUS	AG072852	670 bp DNA linear GSS 03-NOV-2001	
DEFINITION	Pan troglodytes DNA, clone: PTB-064H14.R, genomic survey sequence.		
ACCESSION	AG072852		
VERSION	AG072852.1 GI:16624654		
KEYWORDS	GSS: GSS (genomic survey sequence).		
SOURCE	Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male BAC library clone:PTB-064H14.R.		
ORGANISM	Pan troglodytes		
REFERENCE	Mumukota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.		
AUTHORS	1 (sites)		
TITLE	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Tozoki,Y., Watanabe,H. and Sakaki,Y.		
JOURNAL	BAC end sequences of library PTB		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 670)		
TITLE	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Tozoki,Y., Watanabe,H. and Sakaki,Y.		
JOURNAL	Direct Submission		
COMMENT	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpbes@sc.riken.go.jp, URL:http://ngp.gsc.riken.go.jp/, Tel.:81-45-503-9111, Fax:81-45-503-9170)		
FEATURES	Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.		
PRIMERS			
SEQUENCING:	MJ3rev		
LIBRARY			
VECTOR	: pRS145		
R.Site	: SacI		
R.Site	2 : SacI.		
Location/Qualifiers			
1..670			
/organism=	"Pan troglodytes"		
/db_xref=taxon:9598			
/clone=	"PTB-064H14.R"		
/sex=	"male"		
/cell_type=	"lymphoblast"		
/clone_lib=	"PTB Chimpanzee Male BAC Library"		
BASE COUNT	152 a 190 c 157 g 170 t 1 others		


```

RESULT 9
AGI21669/c
LOCUS AGI21669 689 bp DNA linear GSS 04-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-130M15.F, genomic survey sequence.
ACCESSION AGI21669
VERSION AGI21669.1 GI:16650834
KEYWORDS GSS: GSS (genome survey sequence).
SOURCE Pan troglodytes male lymphoblast DNA, clone: lib-PTB Chimpanzee Male
ORGANISM Pan troglodytes
REFERENCE 1 (sites)
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
Toto, Y., Watanabe, H. and Sakaki, Y.
TITLE BAC end sequences of library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 689)
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
Toto, Y., Watanabe, H. and Sakaki, Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimpanzee@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY Vector : PKS145
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES
source
1. 689
Location/Qualifiers
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-130M15.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 164 a 159 c 176 g 189 t 1 others
ORIGIN
Query Match 33.0%; Score 439.2; DB 12; Length 689;
Best Local Similarity 87.0%; Pred. No. 4, 2e-115;
Matches 507; Conservative 0; Mismatches 66; Indels 10; Gaps 2;

```

```

OY 556 cagagactgctgagcttcctcagagctgaagaatcccaagccttancctgggaagctgac 615
DB 366 CAGAGCTATCTGGATTTCCTAGGCAACTAAGATCCCTTAAGCTTACTGGGAAGGTGAC 307
OY 616 cgcattcattttaaacaatgagcttgcaactttagctcacaccgcaatc----- 667
DB 306 CACATTCATCTTTAAACACAGAGGCTTCGCACTTGTACACCCGAGCAATCAGTAGTA 247
OY 668 -agagagctactaataatgctatcagcagaataagagaggaatgaatgacatcat 726
DB 246 AAGAGGGCTCACTAAATGCTTAATTAGCAAAACAGAGAGTAAAGAAATAGCCAAATAT 187
OY 727 ctattgctctgagagacagagggagaggaagaagattggatataaactcagcattcaag 786
DB 186 TTATTGCTTGAAGCTACAGCGGAGAGACATATGCGGATATATCATGCAATTCAG 127
OY 787 ccagcacaagcacaaccccttgggtccctccatgctatgag 829
DB 126 CCACCAATGGCTACCCCTTTGGGCTCCCTCCCTTGTGTATGG 84
RESULT 10
BH149565/c
LOCUS BH149565 921 bp DNA linear GSS 27-AUG-2001
DEFINITION ENTQ48TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION BH149565
VERSION BH149565.1 GI:15310303
KEYWORDS GSS.
ORGANISM Entamoeba histolytica.
SOURCE Entamoeba histolytica.
ORGANISM Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 921)
AUTHORS Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI: IMSS sheared DNA library (2001)
COMMENT Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 40
High quality sequence stop: 567.
FEATURES
source
1. 921
Location/Qualifiers
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: PHOS1; Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + l method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
BASE COUNT 242 a 216 c 226 g 237 t
ORIGIN

```


Db 82 TTAGCTCACACCCGACCGCATCAGTAGTAAAGAGAGCTCAGTAAATGCTAATATGAGCAA 141

0y 698 aaacaggagggtgaagaacatagccaatctctctcttccctggagagccagcaggaaagacaa 757

Db 142 AAACAGGGTGGTAAAGAAAACAGCCAATCATCTATTCTCCCTGAGAGCAGACAGAGGAGCAA 201

0y 758 ggaatggagataataactaaggcatccaagccagacagcaacacccaccttggctccccc 817

Db 202 TGATGGGATATAAACCCAGGACATTCGAGCCAGCAATGGCTATACCTCTTTGGGCTCCCTC 261

0y 818 ccattgatatggagactctgatttcaactctatctcaactctatuaalcaatgcaactg--ca 875

Db 262 TCTTTGTATGGAGACTCTGT-----TTTCACTATATTAATCTTGGCACTGTACT 311

0y 876 cctcttcggctgcggttttttataatgctcctaagtgagcttttgctgcatccatccacacg 935

Db 312 CTCTCTCTGTGTCATCTTTGTTATGGCTATAACTGACTTTTCACTGCGCATCCACACACTGC 371

0y 936 tggttgcacacgctcacaagaccgcctgctgaactcctcccttggatccagcagaatgctc 995

Db 372 TGTATTGCGCGTGTTCACAGATCCGCCACATGATTTCATCCCTCAGAGATCCGACAGAGTGC 431

0y 996 cactgtgcttcctgcacagcaggaggtaccatctgcaaccccgatcaggctaaagctgtgc 1055

Db 432 CGCTGTGCTCTGATTCAGCGAGGAGCCATTCGACATCCCATCAGGCTAAAGGCTGCT 491

0y 1056 catgttccctgcatactgcaagctgcctgggtgtgtgtcctcctaataagaactgcaacatgtc 1115

Db 492 AATTGTCTCTGCATGGCTAAGTGCCAGAGTGGTCTCTAATGAGCTGAACGCTAGTCACT 551

0y 1116 gggttccatgtgtctctctccatagaccacacggctcttaataagacataaacctacgcga 1175

Db 552 GGGTTCACAGGACTCTCTCGTAGCCACCGGCTTCTAATAGAGCTATAACACTCAGCAC 611

RESULT 14

AG086046/c

LOCUS AG086046 736 bp DNA linear GSS 03-NOV-2001

DEFINITION Pan troglodytes DNA, clone: PTB-084E11.R, genomic survey sequence.

ACCESSION AG086046

VERSION AG086046.1 GI:16637848

KEYWORDS GSS: GSS (genome survey sequence).

SOURCE Pan troglodytes male lymphoblast DNA, clone_1lb:PTB Chimpanzee male BAC library clone:PTB-084E11.R.

ORGANISM Pan troglodytes

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Tojoki, Y., Watanabe, H. and Sakaki, Y.

TITLE BAC end sequences of library PTB

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 736)

AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T., Tojoki, Y., Watanabe, H. and Sakaki, Y.

TITLE Direct Submission

JOURNAL Submitted (02-AUG-2001) Aseo Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 220-0045, Japan (E-mail:chimpbes@sc.riken.go.jp, URL:htp://hnp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)

COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.

PRIMERS

Sequencing: MJ3rev

LIBRARY

Vector : pRS145

R.Site 1 : SacI

R.Site 2 : SacI.

Location/Qualifiers

1..736

/organism="Pan troglodytes"

FEATURES

source

Query Match	Best Local Similarity	Score	DB	Length
Matches 522: Conservative	0: Mismatch	107: Indels	9: Gaps	2:
Query Match	32.3%	Score 429.4	DB 12	Length 736
Best Local Similarity	81.8%	Pred. No. 2.9e-112		
Matches 522: Conservative	0: Mismatch	107: Indels	9: Gaps	2:
Query	670	agaagctcaataatgcatcgaacgcaaaacaggaaggttaagcaatagccatcatccta	729	
Dn	729	AGGGCTCAGTAAATGCTAAT-TGGCGAGAACAGACAGGTAAAGAAATACCACTTCACTA	671	
Qy	730	ttgcctgagacacacgaggaaggaagaaatggaataaactcaagcattcaagca	789	
Dn	670	TGCGCTGAGAGACAGTGGAGGACAAATACGGGATATTAAGCCACAGCATTCACAACTG	611	
Qy	790	gcaaaagaccccccttgggtccccctccatctgtaagagagcctgtcttccatctat	849	
Dn	610	ACAGGCGCTACCCCTGTTGGGTCCCTCCCTTTGTAATGGAGACTGTGTTTTCACATGAT	551	
Qy	850	tcaactataaataatcatgcaactgcatcttctgtgccgtgttttataagctcaagctg	909	
Dn	550	AGGTTTCAAGTGA-----CACATCATGTGTCGCTGTTTATAGGCTCGAGCTG	499	
Qy	910	agctttgttgccatccaccactgtctgttggcaccgtaacagcccgctgtgacttc	969	
Dn	498	AGCTTTCCTCGCATCCACCACTGCTTTTGCGCCCTGTCAACAGACCCGCGCATACATTC	439	
Qy	970	catcccttggatccagcagaagtgtccactgtgtccctcgtatcccaagcagaagtaaccattgc	1029	
Dn	438	CACCCCTCCGGATCCAGCAGGGGTGTGTGCTGTGCTGTGCTGTGATCCAGGAGAGGCCCATTCG	379	
Qy	1030	cactcccgatcaagcattaagctgtgcatctgtctctgcatgctaaagtgcctggttctgt	1089	
Dn	378	GGCTCTCCATCAGGGGTGAAGGCTGGCCCTTTGTTCTTGACAGGATTAAGTCCCGGGTTTGT	319	
Qy	1090	cctaataagacgaacactgtgtcactgtgtgtccatggtgtctctcattcattgacccagagctt	1149	
Dn	318	CGTAATTAGCTGAACACTAATGTCGTGGGTTCCACGAGATCTTCCATGACCCACGCGCTT	259	
Qy	1150	cctaataagcattaaactcaccgcatcgcgcatgagcccaagattccattcctgtgtaactgtgag	1209	
Dn	258	CTAATAGATCTATACACTACACCCGATGGCTCAATATCCATTCCTTGGAAATCCCTGAAG	199	
Qy	1210	ccaagaaccccaagtcacaagaanaangaaagctgtgcacacatttggaaagtgcacactgac	1269	
Dn	198	CCAAAGACTCCAGGTCAGAGAACAAAGAGGCTTGCCCTCATCTGTGGAAGTGGCTCCGACCC	139	
Qy	1270	atttctgtagcggccacacacatcttggagagctgtg	1307	
Dn	138	ATCTCGAAGCGGCGCTGCCACTCTCTGTGGAGCTCTCG	101	
RESULT 15				
A0238712		714 bp	DNA	linear
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				

TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
COMMENT Other GSSs: RPC11-64L19.TK
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamast@tigr.org

Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.bufileo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.bufileo.edu/ordering>) or from Research Genetics (<http://inforesgen.com>). BAC end search page: http://www.tigr.org/tcdb/humgen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1. 714
/organism="Homo sapiens"
/db_xref="GDB:7524474"
/db_xref="taxon:9606"
/clone="RPC1-11-64L19"
/clone_lib="RPC1-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPC11 Human Male BAC Library"

BASE COUNT 183 a 211 c 154 g 166 t
ORIGIN

Query Match 32.2%; Score 427.6; DB 12; Length 714;
Best Local Similarity 83.0%; Freq. No. 9.4e-112;
Matches 565; Conservative 0; Mismatches 92; Indels 24; Gaps 6;

QY 589 atccnaagccttact-gggaagtgacccatccatcttaaacatggggcttgcact 647
DB 1 ATTCTTAAGCTTAAGCTTGGGAGGTGACACACCCTCTTTAAACACAGAGCTTGAAT 60
QY 648 tagctcacaccgacacatc-----agagagctcactaaatgttaatcagc-aa 697
DB 61 CAGCTCACACCCGACCTAGTAAAGAGAGCTCACTAAATATACCAATTAGGCTAA 120
QY 698 aacacagaggttaagaacat-agccaatcatctatgtcctgaagacagcggaagaca 756
DB 121 AAACAGAGGTAAAGAAATATCAATCATATCGCTGAGACACAGGGGAGGACA 180
QY 757 aggatggatataaactcagagcttcaacccaagcaaccccttgggtccct 816
DB 181 ATGATCGGATATTAACCCAGGCAATTGAGCCAGATCAGTAACCTCTTTGGGTCCCT 240
QY 817 cccattgtagggagctctgtttcactctatctactatataatcatgcaactg--c 874
DB 241 CACACTGTATGGAGCTCTGT-----TTTCACTATATTAATCTTGCAACTGCAC 290
QY 875 acctcttggtcgtgtttttatgtcacaagctgtgcttggccatccacactg 934
DB 291 ACTCTTCTGGTCAATGTTTTCGGCTCAAGCTGAGCTTTGCTCCGCTCCACCACTG 350
QY 935 ctgttgcacacgctcaagacccgctgtcactcactcccttggatccagagagtg 994
DB 351 CTGAATGCCGCAATGCAAGACCTGCTTGAATCCACCCCTCCGGATCCGGCAGAGTGT 410
QY 995 ccaactgtctcctgatacagcagagtaaccattgcaactcccgatacaggttaagctg 1054
DB 411 CCGCTGCATCTCTGATCAGAGGAGGACCATTCCTCCGATCAGGCTAAAGGCTTG 470
QY 1055 ccaattgtcctgatacaggttaagctggtttgtcctaataagaactgaacactgtgtcac 1114
DB 471 CCATTTCTCTGCACACAGTAAAGTCTGGGTCTCACTAATCAGGCTGAAACACTGTTCCG 530

QY 1115 tgggtccatggtctctctccatagaccacgcttctaataagactataaacactaccgc 1174
DB 531 TGGGTTCCACAGGTTCTCTTCATGACACAGCTTCTAATAGAGCTAATACATTACAC 590
QY 1175 atggcccaagattccatctctgtgtatctgtgagccaagaacccaggtcagagaagt 1234
DB 591 ATGGCCCAAGGTTCCATTCCTTGGAATCCATGATGCCAAGAACCCAGTCAAGAAATA 650
QY 1235 gaggttcacacacattggga 1255
DB 651 AAGGCCGCCCCCATCTTGGGA 671

Search completed: June 20, 2002, 08:18:47
Job time: 7373 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 09:38:01 ; Search time 97.7 Seconds
(without alignments)
3341.320 Million cell updates/sec

Title: US-09-319-156A-12
Perfect score: 1329
Sequence: 1 tcaaatcgaagagcttag.....gcaagatccccagtaaca 1329

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/pdata/1/lna/5A.COMB.seq:**
2: /cgn2_6/pdata/1/lna/5B.COMB.seq:**
3: /cgn2_6/pdata/1/lna/6A.COMB.seq:**
4: /cgn2_6/pdata/1/lna/6B.COMB.seq:**
5: /cgn2_6/pdata/1/lna/CTUS.COMB.seq:**
6: /cgn2_6/pdata/1/lna/Backfile1.seq:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	731.8	55.1	2946	4	US-09-175-928-3
2	208	15.7	279	1	US-08-686-878A-50
3	208	15.7	279	1	US-08-721-489-4
4	166	12.5	1722	3	US-08-691-563C-58
5	165	12.4	1859	3	US-08-691-563C-46
6	161.8	12.2	80246	4	US-09-078-294-3
7	151.8	11.4	80595	4	US-09-078-294-3
8	80.6	6.1	1704	1	US-08-007-282B-1
9	80.6	6.1	2518	4	US-09-011-745-1
10	80.6	6.1	5855	4	US-09-011-745-1
11	57	4.3	7218	1	US-08-232-463-14
12	50.8	3.8	8060	4	US-08-766-528-1
13	49.2	3.7	8132	4	US-08-766-528-3
14	46.4	3.5	1911	1	US-08-258-420-8
15	46.4	3.5	1965	1	US-08-258-420-9
16	46.4	3.5	3925	4	US-09-011-745-9
17	46.4	3.5	8202	1	US-08-258-420-13
18	44.8	3.4	6028	4	US-09-011-745-5
19	44.6	3.4	1989	4	US-09-232-278A-5
20	44.6	3.4	2499	1	US-08-105-483-310
21	44.6	3.4	2499	1	US-08-709-209-310
22	44.6	3.4	2499	1	US-08-458-101-310
23	44.4	3.3	10970	3	US-08-716-351A-5
24	44.2	3.3	8655	3	US-09-075-272-1
25	43.2	3.3	2462	3	US-09-111-085-1
26	43.2	3.3	7333	4	US-08-766-528-2
27	41.6	3.1	2001	3	US-08-850-961-13

28	41.6	3.1	2001	4	US-09-479-776-13	Sequence 13, Appl
29	41.6	3.1	5109	1	US-08-073-836-1	Sequence 1, Appl
30	41.6	3.1	5176	1	US-08-073-836-2	Sequence 2, Appl
31	41.6	3.1	8332	3	US-08-850-961-1	Sequence 1, Appl
32	41.6	3.1	8332	4	US-09-479-776-1	Sequence 1, Appl
33	41.2	3.1	6312	4	US-09-011-745-7	Sequence 7, Appl
34	40	3.0	1493	4	US-09-376-781-24	Sequence 24, Appl
35	40	3.0	2000	4	US-09-376-781-25	Sequence 25, Appl
36	40	3.0	2000	4	US-09-376-781-30	Sequence 30, Appl
37	40	3.0	3482	3	US-09-111-085-3	Sequence 3, Appl
38	40	3.0	6061	4	US-09-011-745-6	Sequence 6, Appl
39	39	2.9	1914	1	US-08-258-420-7	Sequence 7, Appl
40	39	2.9	8323	1	US-08-110-300A-8	Sequence 8, Appl
41	39	2.9	8323	2	US-08-886-642-8	Sequence 8, Appl
42	39	2.9	8323	4	US-09-433-322B-2	Sequence 2, Appl
43	39	2.9	8323	5	PCT US93-08041-8	Sequence 8, Appl
44	38	2.9	8335	3	US-08-716-351A-1	Sequence 1, Appl
45	37	2.8	4273	3	US-08-795-430-3	Sequence 3, Appl

ALIGNMENTS

```
RESULT 1
US-09-175-928-3
; Sequence 3, Application US/09175928A
; Patent No. 6312921
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavalley, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Mi, Sha
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 6006B A1172A
; CURRENT APPLICATION NUMBER: US/09/175,928A
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2946
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-175-928-3

Query Match      55.1%      Score 731.8: DB 4: Length 2946:
Best Local Similarity 90.7%: Pred. No. 3.7e+240:
Matches 794: Conservative 0; Mismatches 69; Indels 12; Gaps 1;

QY 1 tcaaatcgaagagcttagactgttaaccgcaaaagggggaacctgtatttt 60
    |||||||
DB 2067 tcaaatcgaagagcttagactgttaaccgctgaaggggggaacctgtatttt 2126
    |||||||

QY 61 aggggaagaatgctgttagtatttaactgaatcaactgaagaagtaaga 120
    |||||||
DB 2127 aggggaagaatgctgttagtatttaactgaatcaactgaagaagtaaga 2186
    |||||||

QY 121 aatttagatcggaatataatgtagcagaagagcttcaaaacgaccctggggcct 180
    |||||||
DB 2187 aatttagatcggaatataatgtagcagaagagcttcaaaacgaccctggggcct 2246
    |||||||

QY 181 cctcagccaatgtagtgccttggaactccctctttagagacccctagcagctataatt 240
    |||||||
DB 2247 cctcagccaatgtagtgccttggaactccctctttagagacccctagcagctataatt 2306
    |||||||

QY 241 ttactcctctttagagacccgtatcttcaactcctgttaagtgtgtcttcagaat 300
    |||||||
DB 2307 gctactcctctttagagacccgtatcttcaactcctgttaagtgtgtcttcagaat 2366
    |||||||
```

Oy	301	tgaaagcgtgaagaagctatacaaatagttcttcaaaatggagaaaccccgagatgcagctccatgaacaa	360
Db	2367	cgaaagcgtgaataaacta-----caaatggagcccaagaatgcagctccaaagactaa	2414
Oy	361	aactacacgttgaaaccccttgagaccgsgctcgtctagaactcatgtctctgtatggttaatgaaacttga	420
Db	2415	gactacacgcgaagaccctctgagaccgctcgtcgttagccacgaactctgatatgttaatgaaactcaa	2474
Oy	421	agtcacccctcccgaggaatctcaactgcacaaacccctactacaactccaaltcaatgaatg	480
Db	2475	aggcacacccctcttgaggaaatctcactcgtgcacacacctctactacagccccaattcagccagg	2534
Oy	481	aagcagatgaagaacagctgtgtcagcccaacccctcccaaaagatcttgggtttcttcgtttga	540
Db	2535	aagcagatgttagggcgtctcgtccggccaacccctcccaaaagactttaggtttcttcgtttga	2594
Oy	541	gggtggaacttgaaagacagagactagctcgtgaatttcctctcgtcgtgactaagaatcccaagaact	600
Db	2595	tgggggagctgagagacagagactagctcgtgaatttcctctcgtcgtgactaagaatcccttaagct	2654
Oy	601	aactcgggaaggttgacgcgcatccatctttaaacaatggggctctgcaacttagctcacacccg	660
Db	2655	agctcgggaaggttgacgcacatccacccctttaaacaacggggcttgcaacttagctcacacctg	2714
Oy	661	accacatacgaagagcttcactataatgtctaacgagcaaaaaacgggggttaagaacatagac	720
Db	2715	accacatacgaagagcttcactataatgtctaataatgagcaaaaaacgggggttaagaagaaatagac	2774
Oy	721	aatacatctatgcctctgagacacagacgggaggaagaaagaagatctgggataataaactcagagca	780
Db	2775	aatacatctatgcctctgagacacagacgaaggaaggaacaaatgactcgggataataaaccacaaatc	2834
Oy	781	ttcaagacgaacgaacagcaaacccctctgggtccctcccatctgatatggagactctgtttt	840
Db	2835	ttcagacgcgggaacagcaaacccctctgggtccctccctctgatatggagactctgtttt	2894
Oy	841	caactctatctacactatataatcatgacaacttgcga	875
Db	2895	catgtcattctcaactctataataactctgcgaacttgcga	2929

RESULT 2
 US-08-686-878A-50
 Sequence 50 Application US/0866878A
 Patent No. 5708157
 GENERAL INFORMATION:
 APPLICANT: Jacobs, Kenneth
 APPLICANT: McCoy, John
 APPLICANT: Lavallee, Edward
 APPLICANT: Racie, Lisa
 APPLICANT: Merberg, David
 APPLICANT: Treacy, Maurice
 APPLICANT: Evans, Cheryl
 APPLICANT: Spaulding, Vikki
 TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
 NUMBER OF SEQUENCES: 71
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 Cambridgepark Drive
 CITY: Cambridge
 STATE: Massachusetts
 COUNTRY: U.S.A.
 ZIP: 02140
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/686,878A
 FILING DATE:

```

? CLASSIFICATION: 536
?
? ATTORNEY/AGENT INFORMATION:
?
?   NAME: Brown, Scott A.
?
?   REGISTRATION NUMBER: 32,724
?
?   TELECOMMUNICATION INFORMATION:
?
?   TELEPHONE: (617) 498-8224
?
?   TELEFAX: (617) 876-5851
?
?   INFORMATION FOR SEQ ID NO.: 50:
?
?     SEQUENCE CHARACTERISTICS:
?
?       LENGTH: 279 base pairs
?
?       TYPE: nucleic acid
?
?       STRANDEDNESS: double
?
?       TOPOLOGY: linear
?
?     MOLECULE TYPE: CDNA
?
US-08-686-878A-50

```

Query Match	15.7%	Score 208	DB 1	length 279
Best Local Similarity	85.1%	Pred. No.	2	4e-61
Matches 223	Conservative	1	Mismatches 38	Indels 0
				Gaps 0

QY	614	accgcatccatcctttaacaatgagggtctgcaactgtctacacccgaccaaatacagag	673
Db	1	RCACATCCACCTTTAAACAGCGGGGNTTGCAAAANAAGATTNACCTTACCATACGAG	60
QY	674	ctcaccaaatgctcaatcagcgcaaaacagagaggttaagaacatagccaatcattatgc	733
Db	61	NTCANNTAAATGATNTATNGSCAAAACAGAGCGTAAAGAAATGCGCAATCATCTATTGC	120
QY	734	ctgagagacagcgaggagacaaagatttgggataataactcagcatctcaagcgagcaa	793
Db	121	CTGAGAGCAGCAGGAGGAGCAATGATCGGGATATTAACCAAGTTTGNAGCGCGGCA	180
QY	794	caggaacaccccttgggttccctcccaatgtatgagagctcgttttctaactctatccac	853
Db	181	CGCGAACCCCTTTGGGHCCTCCCTTGATGATGGAGCTGTGTTTCATGCTATTTCAN	240
QY	854	tcatttaatacatgcaactgtca 875	
Db	241	TNTATTAAATNTTGCACATCGA 262	

```

1 RESULT 3
2 US-08-721-489-4
3 Sequence 4, Application US/08721489
4 Patent No. 5786465
5
6 GENERAL INFORMATION:
7 APPLICANT: Jacobs, Kenneth
8 APPLICANT: McCoy, John
9 APPLICANT: Lavallee, Edward
10 APPLICANT: Racle, Lisa
11 APPLICANT: Metberg, David
12 APPLICANT: Treacy, Maurice
13 APPLICANT: Spaulding, Vikki
14 TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
15 TITLE OF INVENTION: ENCODING THEM
16 NUMBER OF SEQUENCES: 5
17
18 CORRESPONDENCE ADDRESS:
19 ADDRESSEE: Genetics Institute, Inc.
20 STREET: 87 CambridgePark Drive
21 CITY: Cambridge
22 STATE: Massachusetts
23 COUNTRY: U.S.A.
24 ZIP: 02140
25
26 COMPUTER READABLE FORM:
27 MEDIUM TYPE: Floppy disk
28 COMPUTER: IBM PC compatible
29 OPERATING SYSTEM: PC-DOS/MS-DOS
30 SOFTWARE: Patent In Release #1.0, Version #1.30
31 CURRENT APPLICATION DATA:
32 APPLICATION NUMBER: US/08/721,489
33 FILING DATE:
34 CLASSIFICATION: 530

```

```

;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: BROWN, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
;
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 279 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; MOLECULE TYPE: cDNA
;
; US-08-721-489-4

```

```

Query Match      15.7%; Score 208; DB 1; Length 279;
Best Local Similarity 85.1%; Pred. No. 2,4e-61;
Matches 223; Conservative 1; Mismatches 38; Indels 0; Gaps 0;

```

```

Db      614  accgcatcatttaaacatgagggttgcaactctcacaccgaccatcagag 673
      1  KCCACATCCCTTTTAACCGGGGNTTGCANANAGATNACCTTGAACCATCAGAG 60
Qy      674  ctactaaatgtctaatcagcaaaacagaggtlaaagaatagccatcatctatgc 733
      61  NTCANTTAATGATNATNTNGCAAAACAGAGGTAAGAATGCCATCATCTATTGC 120
Qy      734  ctgagagcagcagggaagcaagatttgatataaactcaggcaatcaagcagcaa 793
      121  CTGAGACACAGCAGGAGGACATGATCGGATATTAACCAAGTTTNGAGCCGCCAA 180
Db      794  caagcaaccccttgggtccccccttgatagaggagcctgtttcactctatctac 853
      181  CGGCAACCCCTTTGGTCCCTCCCTTTGATGGAGCTNCTTTTCATGCTATTTCAN 240
Qy      854  tctatataatcagcaactgca 875
      241  TNTATTAATNTTGCACACTGCA 262
Db

```

```

RESULT      4
US-08-691-563C-58
; Sequence 58, Application US/08691563C
; Patent No. 6001987
;
; GENERAL INFORMATION:
;
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Glaucia PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND
;
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 92
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: Oliff & Berridge
; STREET: 700 South Washington Street, Suite 300
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22314
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/691,563C
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:

```

```

;
; NAME: Berridge, William P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: MPB 38588
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
;
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1722 base pairs
; TYPE: nucleotide
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; MOLECULE TYPE: cDNA
;
; US-08-691-563C-58

```

```

Query Match      12.5%; Score 166; DB 3; Length 1722;
Best Local Similarity 60.5%; Pred. No. 2e-46;
Matches 333; Conservative 0; Mismatches 200; Indels 17; Gaps 3;

```

```

Qy      2  caaatgagagctttagacttgctaacccgcaaaagagggttaactgtttatttta 61
      547  CAATAACCGCCGAGCCACACCTTCCTGAGAAAGAGAGACTCTGCACCTTCTTA 606
Db      62  gggaagaatgctgtatgtatgtatcgaatcgtgaatcattactgagaagttaagaa 121
      607  GGGGAAGAGTGTCTTTTACACTTACACAGTACGAGGATAGACAGATGCCACTGGCAT 666
Qy      122  attgagatcgaatataatgtagacagagagacctcaaaacactgcaacctgggagctc 181
      667  TTACAGAAAGGGCTTGATATACAGAAATGCTTTCAAACTCTTATACCA-----C 720
Db      182  ctacgcaatgagatgacctgagacctccctctcttagagacctcagagctatataatt 241
      721  CTCTGAGATGGGCAACATGAGCTTCTTCCATTTCTAGCTCCATGCGACCATCTTGCTG 780
Qy      242  ttaactcctcttgagacctgatactcaacttctcttgtaagttgtctcttcagaatt 301
      781  TTACTCACTTTGGGCCCTGTATTTTAAGCTTCTTGCAATTTGTTTCCCTAGAGATC 840
Db      302  gaagctgtaagctacaaatagttcttcaaatggaacccagatgagctcagaactaaa 361
      841  GAACCCATCAAGCTACAGATGCTCTTACAAATGGAACCCCAAGAGTTCAACTAACAAAC 900
Qy      362  atctacccgtgagccctggagccgctgctagactatgctgtatgttaagaaatgaa 421
      901  TTCTACCAAGGAGCCCTGGAACGATCCACTGGAACCTTCCACT-----ACCTAGA 950
Db      422  gtcaaccctccggaggaatctcaactgcaacacccctactacactccaattcagtagga 481
      951  GATTCCTCTGTGAAGACACTACACTGCAAGGCCCTTTTGGCCCTATCCAGCAGAGA 1010
Qy      482  agcagttagagcattgtcagccaaacctcccaacagtaactgttggttctctgttagag 541
      1011  AGTGTAGTAGAGCGGTCACTGCGCAA-ATTCCACAGAGAGTTGGGAGTGTCTGTTAGAG 1069
Db
Qy      542  ggtgactgca 551
      1070  GGGGAGATTGA 1079
Db

```

```

RESULT      5
US-08-691-563C-46
; Sequence 46, Application US/08691563C
; Patent No. 6001987
;
; GENERAL INFORMATION:
;
; APPLICANT: Herve PERRON
; APPLICANT: Frederic BESEME
; APPLICANT: Glaucia PARANHOS-BACCALA
; APPLICANT: Florence KOMURIAN-PRADEL
; APPLICANT: Colette JOLIVET
; APPLICANT: Bernard MANDRAND

```

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
TITLE OF INVENTION: THERAPEUTIC PURPOSES
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: Oliff & Berridge
STREET: 700 South Washington Street, Suite 300
CITY: Alexandria
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22314
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/691,563C
FILING DATE: 02-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Berridge, William P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: WPB 38588
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 1859 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
us-08-691-563C-46

Query Match 12.4% Score 165; DB 3; Length 1859;
Best Local Similarity 58.3%; Pred. No. 4.6e-46;
Matches 369; Conservative 0; Mismatches 247; Indels 17; Gaps 4;

Db 2 caaatcgaagagcttgaactgtctaaacgcaagaggggagacgtttatttta 61
1137 CAAACCGTCAGAGCGCTGAGCTCTCTACCTGCGAAGAGGAGCTGCACTTTCTTA 1256
62 ggggaagaatcgtttagttagttagttagttagttagttagttagttagttagt 121
1257 AGGGAAGAGTGTGTTGTTTACTTAACCAAGTCAGGAGTAGTA--TGAGATGCTGCCGGC 1314
122 atttgaatcgaataataatgtagagcagaagaccccaaacacgtcacccctgggctc 181
1315 ATTTACAGAAAAAGCTTCTGAAATCAGAACAGCGCTTTCAAAT--CCATATACCAAC 1370
182 ctacgaacatgagatgacgtgacgtcctcctcctcctcctcctcctcctcctcctc 241
1371 CTCTGAGATTGGGCAACATGTTTCTTCCCTTTATGTCCTCATGGCCATCTTGTCTA 1430
242 ttactcctcttggacccttgatctcctcaactcctctgttaagtgtctcctcagaatc 301
1431 TTACTCGCCTTTGGGCCCGTATTTTAACTCTGTCATTAATTTGTTTCTTCTAGATC 1490
302 gaagctgtgaagaatcaatagttcttcaaatgaaacccacagatcagtcactgaataa 361
1491 GAGGCATCAACCTACATGATGTTTACAAATGAAACCCCAATAGAGCTCAACTATCAAC 1550
362 atctacgttgagacccctggaccggtcctgtagacatgctcgtgagtttaagtacaa 421
1551 TTCTACTAGAGACCCCTTAGACCAACCCCTGCGCTTTCACCTGCGCTTAAGGTT----- 1605
422 gtacacccctcccgagaaatctcaactgcaaacacccctactcaactcaaatctaga 481
1606 ----CCCTCTGTGAGAGACACTACACATGACGAGGCCCATCTTGGCCCTATCAGAGA 1661
482 agcagttgagcagttgtcagcgaacccctcccaacagtaacttggttctcctgttgag 541

1137 CAAACCGTCAGAGCGCTGAGCTCTCTACCTGCGAAGAGGAGCTGCACTTTCTTA 1256
62 ggggaagaatcgtttagttagttagttagttagttagttagttagttagttagt 121
1257 AGGGAAGAGTGTGTTGTTTACTTAACCAAGTCAGGAGTAGTA--TGAGATGCTGCCGGC 1314
122 atttgaatcgaataataatgtagagcagaagaccccaaacacgtcacccctgggctc 181
1315 ATTTACAGAAAAAGCTTCTGAAATCAGAACAGCGCTTTCAAAT--CCATATACCAAC 1370
182 ctacgaacatgagatgacgtgacgtcctcctcctcctcctcctcctcctcctcctc 241
1371 CTCTGAGATTGGGCAACATGTTTCTTCCCTTTATGTCCTCATGGCCATCTTGTCTA 1430
242 ttactcctcttggacccttgatctcctcaactcctctgttaagtgtctcctcagaatc 301
1431 TTACTCGCCTTTGGGCCCGTATTTTAACTCTGTCATTAATTTGTTTCTTCTAGATC 1490
302 gaagctgtgaagaatcaatagttcttcaaatgaaacccacagatcagtcactgaataa 361
1491 GAGGCATCAACCTACATGATGTTTACAAATGAAACCCCAATAGAGCTCAACTATCAAC 1550
362 atctacgttgagacccctggaccggtcctgtagacatgctcgtgagtttaagtacaa 421
1551 TTCTACTAGAGACCCCTTAGACCAACCCCTGCGCTTTCACCTGCGCTTAAGGTT----- 1605
422 gtacacccctcccgagaaatctcaactgcaaacacccctactcaactcaaatctaga 481
1606 ----CCCTCTGTGAGAGACACTACACATGACGAGGCCCATCTTGGCCCTATCAGAGA 1661
482 agcagttgagcagttgtcagcgaacccctcccaacagtaacttggttctcctgttgag 541

Query Match 12.2% Score 161.8; DB 4; Length 80246;
Best Local Similarity 58.7%; Pred. No. 7.2e-44;
Matches 360; Conservative 0; Mismatches 234; Indels 19; Gaps 4;

Db 2 caaatcgaagagcttgaactgtctaaacgcaagaggggagacgtttatttta 61
58640 caaacgcgtgagcctgtagactctcctcctcctcctcctcctcctcctcctcctc 58698
62 ggggaagaatcgtttagttagttagttagttagttagttagttagttagttagt 121
58699 ggggtagagtggttcttatacctaacagcagtaggaataatgagataccaacccag 58758
122 atttgaatcgaataataatgtagagcagaagaccccaaacacgtcacccctgggctc 181
58759 ttacagaaagagctctcgaatcagacatgacgtcttcaactc-----tatccaac 58812
182 ctacgaacatgagatgacgtgacgtcctcctcctcctcctcctcctcctcctcctc 241
58813 ccttgagttgggagacatggtctcctcctcctcctcctcctcctcctcctcctc 58872
242 ttactcctcttggacccttgatctcctcaactcctctgttaagtgtctcctcagaatc 301
58873 atagtcgaatttggccctgatttcttaacccctcgtgtaaatgttctcctcagagac 58932
302 gaagctgtgaagaatcaatagttcttcaaatgaaacccacagatcagtcactgaataa 361
58933 gaggcacatcagctcagatgtagtcttaacaaatgtaaccccaatgagtcacacaaac 58992
362 atctacgttgagacccctggaccggtcctgtagacatgctcgtgagtttaagtacaa 421
58993 ttctgtgagagacccctgagacccgctggccctc-----atggcctaaag 59043
422 gtacacccctcccgagaaatctcaactgcaaacacccctactcaactcaaatctaga 481
59044 agctccctcctggagaaatctcaactgcaaacacccctcctcctcctcctcctcctc 59103
482 agcagttgagcagttgtcagcgaacccctcccaacagtaacttggttctcctgttgag 541
59104 agtagctagcagtggtcagcgaacccctcccaacagtaacttggttctcctgttgag 59160

Oy	542	ggttcacgcgaagaaaggactacgcggatttccctagcgtctaagaatcccnaagccca	601
Db	59161	ggggcgattcgagaggggaagccaagcctcgsgctcttggtcagrgtgggaccttggagacctt	59220
Oy	602	nctggggaagtctga	614
Db	59221	tgtgctactagctcaa	59233

RESULT 7
 US-09-078-294-3
 : Sequence 3, Application US/09078294
 : Patent No. 6265211
 :
 : GENERAL INFORMATION:
 : APPLICANT: Choo, Kong-Hong Andy
 : APPLICANT: Du Sart, Desiree
 : APPLICANT: Cancilla, Michael R.
 : TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
 : FILE REFERENCE: Dayles Col
 : CURRENT APPLICATION NUMBER: US/09/078,294
 : CURRENT FILING DATE: 1998-05-13
 : NUMBER OF SEQ ID NOS: 29
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO 3
 : LENGTH: 80595
 : TYPE: DNA
 : ORGANISM: Nucleotide sequence of HC-contig
 : US-09-078-294-3

Query Match	11.4%;	Score 151.8;	DB 4;	Length 80595;
Best Local Similarity	58.5%;	Pred. No. 1.9e-40;		
Matches 327;	Conservative	0;	Mismatches 214;	Indels 18;
				Gaps 3;

[illegible]

RESULT 8
US-08-007-282B-1
; Sequence 1, Application US/08007282B

APPLICANT: NAZERIAN, KEYVAN
APPLICANT: CALVERT, JAY G.
APPLICANT: WITTER, RICHARD L.
APPLICANT: YAMAGIDA, NOBORU
TITLE OF INVENTION: VACCINE COMPRISING FOWLPOX VIRUS
TITLE OF INVENTION: RECOMBINANTS EXPRESSING THE ENVELOPE GLYCOPROTEIN OF AN
TITLE OF INVENTION: AVIAN RETICULOENDOTHELIOSIS RETROVIRUS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOIASCH & BIRCH
STREET: 8110 Gatehouse Road Suite 500 East
CITY: Falls Church
STATE: VA
COUNTRY: USA
ZIP: 22042
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/007,282B
FILING DATE: 19930121
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: MURPHY JR., GERALD M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1644-104P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1704 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYDROTHERICAL: NO
ANTI-SENSE: NO
US-08-007-282B-1

Query Match	6.1%	Score 80.6	DB 1	Length 1704
Best Local Similarity 53.7%		Pred. No. 3.6e-17		
Matches 167	Conservative	0	Mismatches 144	Indels 0
QY	2 caaaatcgaagagccttagactgtcctaaccgcaaaaaggagggaacctgtattttta 61			
Db	1387 CAAATATAGAGAGGGGTTACACTATTTGACTGCGGACACAGAGGATATGTCTCGACATC 1446			
QY	62 ggggagaagatgctgttagtagttaaactaatctbgaactaacbagaagaatlaaaga 121			
Db	1447 CAGGAGAACTGTTGTTTACGTAAACAAGTGGGGTATGTAGCTACAGATCCCAAAA 1506			
QY	122 attgagatcgatatatagtagagagagagccttcaaaacacgtgacccctggggccct 181			
Db	1507 CTCACAGAGGACCTTATCCAGAGAAAACGTCACTTACGACACACCCCTGTGGAGCGG 1566			
QY	182 ctcaagccaatggatgcccctggaactctccctctttagagactccttagaagctataatatct 241			
Db	1567 TTGAACGGCTTCTCTCCATATTTGCTACACCTGTGTTGAGCCCTCTGTTGGGCTCATATTG 1626			
QY	242 ttaccctcttggagcccgatatcttcaactctccctgttaagtgttctcttccagaatt 301			

Dd	167	TTCTGACCTCGGGCCGTCATTATGAAGACCCGACTCGATTATACATGACAAATT	1686
Qy	302	gaagctgtaaa	312
Dd	1687	CAGGCAGTAAA	1697

```

RESULT 9
US-09-011-745-1
: Sequence 1, Application US/09011745
: Patent No. 6165715
: GENERAL INFORMATION:
: APPLICANT: Collins, Mary KL
: APPLICANT: Weiss, Robin A
: APPLICANT: Takeuchi, Yasuhito
: APPLICANT: Cosseil, Francis-Lolc
: TITLE OF INVENTION: Transposon systems
: FILE REFERENCE: 09/011,745
: CURRENT APPLICATION NUMBER: US/09/011,745
: CURRENT FILING DATE: 1996-06-22
: EARLIER APPLICATION NUMBER: PCT/GB96/02061
: EARLIER FILING DATE: 1996-08-23
: EARLIER APPLICATION NUMBER: GB9517263.1
: EARLIER FILING DATE: 1995-08-23
: NUMBER OF SEQ ID NOS: 29
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 2518
: TYPE: DNA
: ORGANISM: RD114
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)
: OTHER INFORMATION: n is any nucleotide
US-09-011-745-1

```

	Query Match	6.1%	Score 80.6:	DB 4	Length 2518:
	Best Local Similarity	53.7%	Pred. Num. 4.7e-17:		
	Matches 167:	Conservative	0:	Mismatches 144:	Indels 0:
					Gaps 0:
QY	2 caaaatcgaagagcctttagacttgcactgtaaccgcaaaagagggggaaccggttatattta 61				
Db	1930 caaaatagaggagagcctgagcacttaactaaacgagaagaaggggaattgttttagcctta 1989				
QY	62 ggggagaagaatcgtctgttagtagtataatcaatctggaatcattactggaagaatttaagaa 121				
Db	1990 caaaggaaaatctcgttttttatbgtctaacaaagttaggaatgttgaagaacaaataagaaacc 2049				
QY	122 atttgagatcgaaatatagttagagcagagagccttcaaaagacatgcaccctgggggcc 181				
Db	2050 ctacagaagaatctcaaaaaaacgaggaagagcggcgaaccaaccctctcggagccgg 2109				
QY	182 ctcaagcaaatgtagtccctcctgagactctccctcttcttagagaccccttagcagctataatatt 241				
Db	2110 ctgcaaggccttctcttcgttaacctctactctctccgcgggccccttaaccctctacc 2169				
QY	242 ttaactcctctttagagccctgtaactctcaactcctttagaattgtgctctccagaatt 301				
Db	2170 atactaaccaatctgggcagatcgcttttcagtcgcctccatctgacctcaattatgatagactt 2229				
QY	302 gaagcctgtataa 312				
Db	2230 aatgtttgtaca 2240				

RESULT 10
US-09-011-745-8
: Sequence 8, Application US/09011745
: Patent No. 6165715
: GENERAL INFORMATION:
: APPLICANT: Collins, Mary KL
: APPLICANT: Weiss, Robin A

```

1  APPLICANT: Takeuchi, Yasuhiro
2  APPLICANT: Cosset, Francois-Joic
3  TITLE OF INVENTION: Expression systems
4  FILE REFERENCE: 09/011,745
5  CURRENT APPLICATION NUMBER: US/09/011,745
6  CURRENT FILING DATE: 1998-06-22
7  EARLIER APPLICATION NUMBER: PCT/GB96/02061
8  EARLIER FILING DATE: 1996-08-23
9  EARLIER APPLICATION NUMBER: GB9511263.1
10 EARLIER FILING DATE: 1995-08-23
11 NUMBER OF SEQ ID NOS: 29
12 SOFTWARE: PatentIn Ver. 2.0
13
14 SEQ ID NO 8
15
16 LENGTH: 5865
17
18 TYPE: DNA
19 ORGANISM: Artificial Sequence
20
21 FEATURE:
22 OTHER INFORMATION: Description of Artificial Sequence: Portion of
23 FEATURE:
24 OTHER INFORMATION: construct
25
26 FEATURE:
27 NAME/KEY: misc.feature
28 LOCATION: (3611)
29 OTHER INFORMATION: n is any nucleotide
30
31 FEATURE:
32 NAME/KEY: misc.feature
33 LOCATION: (3612)
34 OTHER INFORMATION: n is any nucleotide
35
36 FEATURE:
37 NAME/KEY: misc.feature
38 LOCATION: (3613)
39 OTHER INFORMATION: n is any nucleotide
40
41 FEATURE:
42 NAME/KEY: misc.feature
43 LOCATION: (3614)
44 OTHER INFORMATION: n is any nucleotide
45
46 FEATURE:
47 NAME/KEY: misc.feature
48 LOCATION: (3799)
49 OTHER INFORMATION: n is any nucleotide
50
51 FEATURE:
52 NAME/KEY: misc.feature
53 LOCATION: (3800)
54 OTHER INFORMATION: n is any nucleotide
55
56 FEATURE:
57 NAME/KEY: misc.feature
58 LOCATION: (3801)
59 OTHER INFORMATION: n is any nucleotide
60
61 FEATURE:
62 NAME/KEY: misc.feature
63 LOCATION: (3802)
64 OTHER INFORMATION: n is any nucleotide
65
66 US-09-011-745-8

```

	Query Match	Similarity	6.1%	Score	80.6	DB 4	Length	5865
	Best Local	Similarity	53.7%	Pred.	No.	8.3e-17		
	Matches	167	Conservative	0	Mismatches	144	Indels	0
							Gaps	0
Oy	2	caaaatcgagaagccttagtactgtctaaacgcgaagaagggggaacctgtattttha	61					
Db	2788	caaatatgagggagctgactgactactaaacgcgagaagaaggagattgttttagccta	2847					
Oy	62	ggggagaagatcgtctgtatgatgtataatcaatcttggaatcattaccggaagaattaaaga	121					
Db	2848	caaggaaaatgcgttttattatgtctaaacgattgaagatgttgaagaacaaataagaacc	2907					
Oy	122	atttgatctgataatagttagagcagagacgcctcaaacacgcgcaccctggggccctc	181					
Db	2908	ctacagaagaattcaaaaacgcgggaagaacccgcgcacccaacctctctgcgcggcg	2967					
Oy	132	ctcagccaatgtagtccctcgaaactctccctcttaggaacctctgacagctataatatt	241					
Db	2968	ctgcagagcttctcttcgcgtacctactactctccctcggaagccctctcaacctcttacc	3027					

Db 1679 AAGGAGGATGCTGTTTATGTCATTCATGAGGCGCATCAGAGCTCCATGAACAAA 1738
Oy 122 attgaagtcgaataatgtagagcagagaccctcaaacacgcccctgg 175
Db 1739 CTTAGAGAAAGCTTGACAGAGCTCGAAGGAAAGAACTACTCAAGGGTGG 1792

RESULT 13
US-08-766-528-3

; Sequence 3, Application US/08766528

; Patent No. 6190861

; GENERAL INFORMATION:

; APPLICANT: Jay A. Fishman

; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS

; NUMBER OF SEQUENCES: 74

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: LAHIVE & COCKFIELD, LLP

; STREET: 60 State Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02109-1875

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; FILING DATE: US/08/766,528

; PRIORITY DATE:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/572,645

; FILING DATE: 14-DEC-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Louis Myers

; REGISTRATION NUMBER: 35,965

; REFERENCE/DOCKET NUMBER: MGP-038CP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617)227-7400

; TELEFAX: (617)227-5941

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 8132 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; -08-766-528-3

Query Match 3.7%; Score 49.2; DB 4; Length 8132;
Best Local Similarity 55.2%; Pred. No. 6e-06; Mismatches 78; Indels 0; Gaps 0;

Oy 2 caaatcgaaagagcttgaacttgctaacgcgcaaaagagggggaacctgttatttla 61
Db 7153 CAGATTAAGAGAGGCTGATTATTATTCTAAGAGAGGAGATATGTGTACCTTG 7212
Oy 62 ggggaagatgctgttagatgtaataatcaatctggaatcattacagaaagttaaaaa 121
Db 7213 AAGGAGGAATGCTGTTTATGTCATTCAGGGGCCATCAGAGCTCCATGAAG 7272
Oy 122 attgaagtcgaataatgtagagcagagaccctcaaacacgcccctgg 175
Db 7273 CTTAGAGAAAGCTTGACAGAGCTCGAAGGAAAGAACTACTCAAGGGTGG 7326

RESULT 14
US-08-258-420-8
; Sequence 8, Application US/08258420
; Patent No. 5710037

; GENERAL INFORMATION:
; APPLICANT: Nienhuis, Arthur W.
; APPLICANT: Vanlo, Eljo F.
; TITLE OF INVENTION: No. 5710037e1 Retroviral Envelope and LTR and Retroviral Ve
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Carrella, Byrne, Bain, Giffillan, Cecchi, Stewart & Olstein

; STREET: 6 Becker Farm Road

; CITY: Roseland

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07068

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch diskette

; COMPUTER: IBM PS/2

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: Wordperfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/258,420

; FILING DATE: 10-JUN-1994

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Olstein, Elliot M.

; REGISTRATION NUMBER: 24,025

; REFERENCE/DOCKET NUMBER: 271010-208

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201-994-1700

; TELEFAX: 201-994-1744

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1911 bases

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: genomic DNA

; FEATURE:

; NAME/KEY: retroviral envelope sequence

; US-08-258-420-8

Query Match 3.5%; Score 46.4; DB 1; Length 1911;
Best Local Similarity 48.1%; Pred. No. 2.1e-05;
Matches 177; Conservative 0; Mismatches 176; Indels 15; Gaps 1;

Oy 2 caaatcgaaagagcttgaacttgctaacgcgcaaaagagggggaacctgttatttla 61
Db 1528 CAGAACCCGAGAGGCGCTGATTGCTATTCCTAAGAGAGGAGCTCTCGCGCCCTA 1587
Oy 62 ggggaagatgctgttagatgtaataatcaatctggaatcattacagaaagttaaaaa 121
Db 1588 AAGAGAGATGCTGTTTATGTCAGACACAGGCGCTAGTGAAGAGAGATGCGCAA 1647
Oy 122 attgaagtcgaataatgtagagcagagaccctcaaacacgcccctggc--- 178
Db 1648 TTAAGAGAAAGCTTAAACAGACAAAAGTATTGAGACAGGCCAAGATGTTGCA 1707
Oy 179 -----ctctgaagcaatgagatgcccctggaactctccctcttgaggacctta 226
Db 1708 GGGCTGTTAATAGATCCCGCTGTTTACACCTTAATCTCACCATATGAGACTCTTA 1767
Oy 227 gaagtcataatcttactcctcttggaacctglatctcaactctctgtaagt 286
Db 1768 ATAGTACTTACTACTTACTTACTCTTGAGCTTCATTCGAATGATGCAATTT 1827
Oy 287 gtcttcagaaatgaaagctgtaagactacaatagttcttcaaaatgaaacccaga 346
Db 1828 GTTAAAGACAGGATATCAGTGTCCAGGCTGAGTTTGTGACTCAACATATACACAGCTG 1887
Oy 347 cagttcat 354

THIS PAGE BLANK (USPTO)

PR 26-NOV-1996; 96US-0756429.
 XX (INMR) BIO MERIEUX.
 XX Bedin F, Beseme F, Jolivet-Reynaud C, Komurian-Pradel F;
 PI Mandrand B, Paranhos-Baccala G, Perron H;
 DR MPI: 1998-322732/28.
 DR P-PSDB: AAM71069.
 PT New nucleic acid from retroviruses - useful for diagnosis,
 prevention and treatment of, e.g. multiple sclerosis
 XX
 PS Disclosure: Pages 187-188; 286pp: English.
 CC The present sequence represents a multiple sclerosis (MS) associated
 CC retrovirus (MSRV) genomic fragment used in the method of the
 CC invention. The invention provides complete or partial genomic
 CC sequences of the MSRV-1 pol gene, gag gene and env gene, and
 CC polypeptides encoded by these genes. The invention also provides
 CC antibodies raised against the polypeptides. The genomic sequences,
 CC polypeptides and antibodies are also claimed useful for diagnosing
 CC infection by MS and rheumatoid arthritis-associated viruses, and also
 CC for prevention and treatment of infection with these viruses.
 XX
 SO Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;

Query Match 99.8%; Score 1326; DB 19; Length 1329;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcaaaatcgaagagcttagacttgcgaacgccaagaaggagggaacgtttattt 60
 Db 1 tcaaaatcgaagagcttagacttgcgaacgccaagaaggagggaacgtttattt 60
 QY 61 agggagaaatgctgtgtatgtatcaatcgtgaatcattactggaagttaaga 120
 Db 61 agggagaaatgctgtgtatgtatcaatcgtgaatcattactggaagttaaga 120
 QY 121 aattggaatcgaatataatgtagagcagaggaacgttcacaaactgcacctgggact 180
 Db 121 aattggaatcgaatataatgtagagcagaggaacgttcacaaactgcacctgggact 180
 QY 181 cctcgaacatgtagctgctggaactcctcctttagaactctagagctataatatt 240
 Db 181 cctcgaacatgtagctgctggaactcctcctttagaactctagagctataatatt 240
 QY 241 tttaactctcttggagccttalcctgaactccttcttgaagtgtgctctccaagat 300
 Db 241 tttaactctcttggagccttalcctgaactccttcttgaagtgtgctctccaagat 300
 QY 301 tgaagctgtaaagctacaaatagctctcaaatggaaccccaagatgacgtacaa 360
 Db 301 tgaagctgtaaagctacaaatagctctcaaatggaaccccaagatgacgtacaa 360
 QY 361 aatctacgttggaacccctgagcggcctgctagactatgctctgtaatgacattga 420
 Db 361 aatctacgttggaacccctgagcggcctgctagactatgctctgtaatgacattga 420
 QY 421 agtcaacccctccgaggaatctcaactgcaacacccctacacacccctcaattcgatgg 480
 Db 421 agtcaacccctccgaggaatctcaactgcaacacccctacacacccctcaattcgatgg 480
 QY 481 aagcagcttagagcaagtctcagcaacacccccaacagtaacttggtcttcctgttga 540
 Db 481 aagcagcttagagcaagtctcagcaacacccccaacagtaacttggtcttcctgttga 540
 QY 541 gggctgactgagagacagagctagctgtattcctaggctgactaagaatcccaagcct 600
 Db 541 gggctgactgagagacagagctagctgtattcctaggctgactaagaatcccaagcct 600
 QY 601 anctgggaagtgacgcacatcctttaaacatgaggcttgcaactagctcaacccg 660

Db 601 anctgggaagtgacgcacatcctttaaacatgaggcttgcaactagctcaacccg 660
 QY 661 accaatacagagctcactaaatgctatcagcgaacaaacagaggttaagcaatagcc 720
 Db 661 accaatacagagctcactaaatgctatcagcgaacaaacagaggttaagcaatagcc 720
 QY 721 aatcattatctgcttgagagcagcagcgggaaggaaagatctggatataaactcagga 780
 Db 721 aatcattatctgcttgagagcagcagcgggaaggaaagatctggatataaactcagga 780
 QY 781 ttcaagcagcaacagcaaccccttgggtccctcccatctgatatggagctcgtttt 840
 Db 781 ttcaagcagcaacagcaaccccttgggtccctcccatctgatatggagctcgtttt 840
 QY 841 cactatttcaactatataatcattgaactgcaactctctctgctcgttttttaagg 900
 Db 841 cactatttcaactatataatcattgaactgcaactctctctgctcgttttttaagg 900
 QY 901 ctcaagctgagctttgttgcgcacatccacactgctgttgcacacgctcacaagaccgt 960
 Db 901 ctcaagctgagctttgttgcgcacatccacactgctgttgcacacgctcacaagaccgt 960
 QY 961 gctgacttccatcccttggatccagcagagtgctcactgtgctcttataccaggaagt 1020
 Db 961 gctgacttccatcccttggatccagcagagtgctcactgtgctcttataccaggaagt 1020
 QY 1021 accaattgcacatcccgatcgaacgcttaagagcttgcaactgtctcgtcgttaagtcc 1080
 Db 1021 accaattgcacatcccgatcgaacgcttaagagcttgcaactgtctcgtcgttaagtcc 1080
 QY 1081 tgggttctcctaataagaaactggaacactgctggaactggttccatggtctcttccatgac 1140
 Db 1081 tgggttctcctaataagaaactggaacactgctggaactggttccatggtctcttccatgac 1140
 QY 1141 ccaaggtcttcaatagactataacactcacccgaatgagcccaagaattcattctctgtga 1200
 Db 1141 ccaaggtcttcaatagactataacactcacccgaatgagcccaagaattcattctctgtga 1200
 QY 1201 tctgtgaagcgaagaaacccaggtgtagaagaagttagagcttgccacacatttgggaagtgg 1260
 Db 1201 tctgtgaagcgaagaaacccaggtgtagaagaagttagagcttgccacacatttgggaagtgg 1260
 QY 1261 cccactgcaatttggtagcggcccaacacatcttgtagagctgtggagcaagatccc 1320
 Db 1261 cccactgcaatttggtagcggcccaacacatcttgtagagctgtggagcaagatccc 1320
 QY 1321 ccagtaaca 1329
 Db 1321 ccagtaaca 1329

RESULT 2
 AAX29704
 ID AAX29704 standard; DNA; 1329 BP.
 XX
 AC AAX29704;
 XX
 DT 08-JUN-1999 (first entry)
 XX
 DE Clone 5M6 from MSRV-1.
 XX
 KW Multiple sclerosis; virus; diagnosis; prophylaxis; therapy; expression;
 KM rheumatoid polyarthritis; ss.
 OS Multiple sclerosis related virus type 1.
 XX
 PN FR2765588-A1.
 XX
 PD 08-JAN-1999.
 XX
 PF 07-JUL-1997; 97FR-0008816.
 XX

PR 07-JUL-1997; 97F0008816.
XX (INMR) BIO MERIEUX.
XX WPI: 1999-098275/09.
DR P-PSDB: AAM9554.
XX
PT Nucleic acid sequences of retrovirus called MSRV-1 - associated with
PT multiple sclerosis or rheumatoid polyarthritis
XX
PS Claim 1; Page 39-40; 83pp; French.
XX
CC This sequence represents clone 5M6 from a novel multiple sclerosis
CC related virus type 1 (MSRV1). The sequence can be used in diagnostic,
CC prophylactic or therapeutic compositions to inhibit expression of a
CC multiple sclerosis related virus and/or virus associated with
CC rheumatoid polyarthritis.
XX
SQ Sequence 1329 BP; 346 A; 352 C; 288 G; 340 T; 3 other;

Query Match 99.8%; Score 1326; DB 20; Length 1329;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 tcaaaatcgagagcttagactgtcctaacgccaagaaggggaacctgtatttt 60
DB 1 tcaaatcgagagcttagactgtcctaacgccaagaaggggaacctgtatttt 60
QY 61 agggagaagaatgctgtatgtatgttaatcaatcttggaatcattactagagaagttaaaga 120
DB 61 agggagaagaatgctgtatgtatgttaatcaatcttggaatcattactagagaagttaaaga 120
QY 121 aatttagatcgataataatgtagagagagaccccttcaaacacgtgacccctggggcct 180
DB 121 aatttagatcgataataatgtagagagagaccccttcaaacacgtgacccctggggcct 180
QY 181 cctcagccaatgtatgcccgtgactccctctcttagagaccccttagacagataatatt 240
DB 181 cctcagccaatgtatgcccgtgactccctctcttagagaccccttagacagataatatt 240
QY 241 ttactctctttagacccctgtatctcaactctctgttaagtgtgtctcttcacgaat 300
DB 241 ttactctctttagacccctgtatctcaactctctgttaagtgtgtctcttcacgaat 300
QY 301 tgaagctgttaagctataaataatttctcaaatggaaccccaatgtagccatgactaa 360
DB 301 tgaagctgttaagctataaataatttctcaaatggaaccccaatgtagccatgactaa 360
QY 361 aatctacgctgagaccccttgagccgctgtctagactatgtctcgatgttaatgacattga 420
DB 361 aatctacgctgagaccccttgagccgctgtctagactatgtctcgatgttaatgacattga 420
QY 421 agtcaacccctccgaggaatctcaactgtcacacccctactactcaactcaattcaatgtag 480
DB 421 agtcaacccctccgaggaatctcaactgtcacacccctactactcaactcaattcaatgtag 480
QY 481 aagcagtttagagcagtgtagcgaacactccccaacagtagtctgggtttctcgtttaga 540
DB 481 aagcagtttagagcagtgtagcgaacactccccaacagtagtctgggtttctcgtttaga 540
QY 541 gggctgagctagagagacagagctagctgattctcttagagctgacataaataatcccaagcct 600
DB 541 gggctgagctagagagacagagctagctgattctcttagagctgacataaataatcccaagcct 600
QY 601 anctgggaaggtgacccgacatcctttaaacaatgagggtctgcaacttagctcacaccg 660
DB 601 anctgggaaggtgacccgacatcctttaaacaatgagggtctgcaacttagctcacaccg 660
QY 661 accaatcagagagctcctcaataatgctcaatcagcaaaaaacaggaagttaaagcataagcc 720
DB 661 accaatcagagagctcctcaataatgctcaatcagcaaaaaacaggaagttaaagcataagcc 720

QY 721 aatcattatgtcctgtagagacagccggaaggaagaattggatataaactcagga 780
DB 721 aatcattatgtcctgtagagacagccggaaggaagaagaattggatataaactcagga 780
QY 781 tcaagccagacaagaacaccccttgggtccctcccatgtatagggagctcgtttt 840
DB 781 tcaagccagacaagaacaccccttgggtccctcccatgtatagggagctcgtttt 840
QY 841 cactctattcactctatataatcaatgacactgacactctctgtcgtgttttttag 900
DB 841 cactctattcactctatataatcaatgacactgacactctctgtcgtgttttttag 900
QY 901 ctaagctgagcttgttgcacatccacacacgtgttggcaccgttcacagaccgct 960
DB 901 ctaagctgagcttgttgcacatccacacacgtgttggcaccgttcacagaccgct 960
QY 961 gctgacttccatcccttggatccagagagtgccaagtgtgctcctgataccagcgaagt 1020
DB 961 gctgacttccatcccttggatccagagagtgccaagtgtgctcctgataccagcgaagt 1020
QY 1021 acccattgcccactccgatacagctaaagcttgccatttctctgataagtgcc 1080
DB 1021 acccattgcccactccgatacagctaaagcttgccatttctctgataagtgcc 1080
QY 1081 tgggttgccttaataagaactggaacactgtgcaactgggttccatgtctctccatgac 1140
DB 1081 tgggttgccttaataagaactggaacactgtgcaactgggttccatgtctctccatgac 1140
QY 1141 ccaagccttctaataagactataaactcaccgcatggtggccaagatccattccttgta 1200
DB 1141 ccaagccttctaataagactataaactcaccgcatggtggccaagatccattccttgta 1200
QY 1201 tctgttagggccaagaaccccgatcagaangtgaagcttgcacatttgggaatgg 1260
DB 1201 tctgttagggccaagaaccccgatcagaangtgaagcttgcacatttgggaatgg 1260
QY 1261 cccactgccatttgtgtagcgccaccacatcttggagctgtggaagcaagatccc 1320
DB 1261 cccactgccatttgtgtagcgccaccacatcttggagctgtggaagcaagatccc 1320
QY 1321 ccagttaaca 1329
DB 1321 ccagttaaca 1329

RESULT 3
AAS68626
ID AAS68626 standard; cDNA; 7466 BP.
XX
AC AAS68626;
XX
DT 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #4430.
XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX

PR 24-FEB-2000; 2000US-0184841.
PR 24-FEB-2000; 2000US-0185213.
PR 24-FEB-2000; 2000US-0185216.
PR 12-MAY-2000; 2000US-0203785.
PR 15-MAY-2000; 2000US-0204226.
PR 16-MAY-2000; 2000US-0204821.
PR 16-MAY-2000; 2000US-0204821.
PR 16-MAY-2000; 2000US-0204908.
PR 16-MAY-2000; 2000US-0205232.
PR 17-MAY-2000; 2000US-0204815.
PR 17-MAY-2000; 2000US-0204863.
PR 17-MAY-2000; 2000US-0205221.
PR 17-MAY-2000; 2000US-0205285.
PR 17-MAY-2000; 2000US-0205286.
PR 17-MAY-2000; 2000US-0205287.
PR 17-MAY-2000; 2000US-0205323.
PR 17-MAY-2000; 2000US-0205324.
XX (INCYTE GENOMICS INC.)

PI Chen SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
PI D'Sa SA, Ansley S, Dahl CR, Dam TC, Daniels SE;
PI Dufour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;
PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockreher TK, Daffo A;
PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
PI Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
XX
XX MPI: 2001-502867/55.
DR P-PSDB: AA019431.

PT Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
PT enzymes, hormones and receptors, useful in diagnostics and therapeutics
XX
XX -
XX Claim 1: page 304; 522pp; English.

XX The invention relates to polynucleotides (I) encoding diagnostic and
CC therapeutic (DITHP) polypeptides (II), which include e.g. enzymes,
CC and proteins involved in growth and development and receptors. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate DITHP expression. For example, (I) and
CC (II) may be used to treat disorders associated with decreased polypeptide
CC expression by rectifying mutations or deletions in a patient's genome,
CC that affect the activity of the DITHPs, by expressing inactive proteins
CC or supplementing the patient's own production of them. (I) and (II)
CC may be used to treat diseases, for example, cell proliferative disorder,
CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
CC leukemia, autoimmune disorders, and respiratory disorders. Additionally,
CC (I) may be used to produce the DITHPs, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the protein. (I) and
CC its complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids in
CC samples, and therefore which patients may be in need of restorative
CC therapy. (II) may also be used as antigens in the production of
CC antibodies against DITHPs and in assays to identify modulators of DITHP
CC expression and activity. The anti-DITHP antibodies and antagonists may
CC also be used to down regulate expression and activity. The anti-DITHP
CC antibodies may also be used as diagnostic agents for detecting the
CC presence of DITHPs in samples (e.g. by enzyme linked immunosorbant
CC assay (ELISA)). AA530986-AA531196 represent human diagnostic and
CC therapeutic (DITHP) polynucleotides of the invention.

XX Sequence 1393 BP; 375 A; 352 C; 304 G; 360 T; 2 other;

Query Match 58.6%; Score 778.8; DB 22; Length 1393;
Best Local Similarity 87.2%; Pred. No. 6,7e-245;
Matches 904; Conservative 0; Mismatches 112; Indels 21; Gaps 4;

OY 302 gaagctgaagcgaacaaatagttcttcaatggaacccagagtcaggtcactgaactaa 361
DB 1 gaagctgaacactaatgtcttccaagaagaccagagtcaggtcactgaactaa 60

OY 362 atcacgctgagccaccctgagccgctctagactaactgctctgatttaattgaactgaa 421
DB 61 attacacagacaccctcgagccgctctgagccacatgacagatgttaattgatatcan 120
OY 422 gtcaaccctccgaggaatctcaactgcaaacaccctactacatcccaattagtaga 481
DB 121 agactcc-cccaaggaattcaactgcaaacaccctactacatcccaattagtaga 179
OY 482 agcagttagagcaggttgtagcgaacactcccaacagtaattggtttctctttagag 541
DB 180 agcagttagagcaggttgtagcgaacactcccaacagtaattggtttctctttagag 229
OY 542 gctgagctgagagacgaagtaactgtagattctctagagctactcaagatcccaagcta 601
DB 240 cgggagctgagagacgaagtaactgtagattctctagagctactcaagatcccaagcta 239
OY 602 nctgggaagtgagccgcatccatttaacaatggggcttgcaactagtcacaccca 661
DB 300 gctgggaagtgagccgcatccatttaacaacgggcttgcaactagtcacaccca 359
OY 662 ccaatc-----agagagctcaactaaatgctaaatcagggcaaaaagaggtagaag 712
DB 360 ccaatcaggtagtaaaagagagctcaactaaatgctaaatcagggcaaaaagaggtagaag 419
OY 713 caatagcgaatcattcttctctgtagagcagcaggaaggaacagagattgagataaa 772
DB 420 aaatagcgaatcattcttctctgtagagcagcaggaaggaaggaacagagattgagataaa 479
OY 773 ctcaagcatcaacgaacgaacgaaccccttggtgctccctccatgtatggagc 832
DB 480 cccaagcatcagagcagcagcagcagccttggtgctccctccatgtatggagc 539
OY 833 tctgtttcactcattctcactcatttaaatcaatgcaacagcactcttcgttcgctgt 892
DB 540 tctgtt-----ttcactcatttaaatcaatgcaacagcactcttcgttcgctgt 589
OY 893 ttctatgctcaagctagacttctgttcgcacatccacacagctgtttgccaaccgtaca 952
DB 590 tgttaagcttgaagctagacttctgttcgcacatccacacagctgtttgccaaccgtaca 649
OY 953 gaaccgctgctgacttccatccctttgtagcagcagagtgctcaactgtgctctgacc 1012
DB 650 gacttgcgctgacttccatccctttgtagcagcagagtgctcaactgtgctctgacc 708
OY 1013 agcaggttaccatgctcactcccgatccgagctaaaggcttgcatgttccctgtaggc 1072
DB 709 agcaggttaccatgctcactcccgatccgagctaaaggcttgcatgttccctgtaggc 768
OY 1073 taagtgcctggtttgtctcaataatagaactgaacactgtgcaactgtgctcaatgttctct 1132
DB 769 taagtgcctggtttgtctcaataatagaactgaacactgtgcaactgtgctcaatgttctct 828
OY 1133 tcatagcccaagcttcaatagactataaactcaacgcatgagcccaagattccatt 1192
DB 829 tccgtgacccaagcttcaatagactataaactcaacgcatgagcccaagattccatt 888
OY 1193 ccttgatctgtaggacgaagaccccaagcttgaggaanngtgggcttgccacattg 1252
DB 889 ccttgatctgtaggacgaagaccccaagcttgaggaanngtgggcttgccacattg 948
OY 1253 ggaagtggcccaactgcatcttggtagcggcccaacacatcttggagagctgtggagca 1312
DB 949 ggaagtggccctgcgcgcacatttggagtgagcttgcacacatcttggagagcttggagca 1008
OY 1313 aggatccccaagtaca 1329
DB 1009 aggaccccttgtaaca 1025

RESULT 5
AAA63826
ID AAA63826 standard; DNA; 2030 BP.
XX

AC AAA63826;
 XX 04-DEC-2000 (first entry)
 DT Nucleotide sequence of the MSRV-1 3' env and LTR regions.
 XX MSRV-1: pol region; long terminal repeat; LTR; RUS region; retrovirus;
 KM ss.
 XX Multiple Sclerosis retrovirus 1.
 OS
 XX
 FH Key Location/Qualifiers
 FT 1..1629
 FT CDS
 FT sig_peptide
 FT 1..81
 FT /note= "Contains one termination codon"
 FT /tag= a
 FT CAAT_signal
 FT 1800..1807
 FT /tag= b
 FT CAAT_signal
 FT 1858..1864
 FT /tag= c
 FT TATA_signal
 FT 1906..1911
 FT /tag= d
 FT polyA_signal
 FT 1996..2002
 FT /tag= e
 FT /tag= f
 XX MO200047745-A1.
 XX 17-AUG-2000.
 XX 15-FEB-2000; 2000WO-1B00159.
 XX 15-FEB-1999; 99EP-0420041.
 XX (INMR) BIO MERIEUX.
 XX Paranhos-Baccala G, Perron H, Komurian-Pradel F;
 DR WPI: 2000-506097/45.
 DR P-PSDB: AAB08195.
 XX Nucleotide fragment of LTR-RUS region from Multiple Sclerosis
 PT retrovirus (MSRV) used to detect the presence of MSRV-1 retrovirus in a
 PT biological sample
 PS
 PS Disclosure: Fig 2; 23pp; English.
 XX The present sequence represents the nucleotide sequence corresponding
 CC to the 3' env region and long terminal repeat sequences from clone
 CC CL6 of Multiple Sclerosis retrovirus (MSRV-1). The specification
 CC describes a long terminal repeat (LTR)-RUS region which encodes the
 CC expression of a MSRV-1 protein. This is unusual for LTRs, in
 CC particular in the RUS region. The sequence includes CAAT and TATA
 CC signals which are present in the U3 and R regions and are not directed
 CC towards the CDS indicated in the features table. Probes and antibodies
 CC to the MSRV-1 retrovirus protein and encoding polynucleotide sequences
 CC are used to detect the presence of MSRV-1 retrovirus in a biological
 CC sample.
 XX
 SO Sequence 2030 BP: 574 A; 559 C; 387 G; 510 T; 0 other;

QY 121 aattgagatcgaatataatgtagagcagaggacccctcaaaacatgcacctggggcct 180
 Db 1260 aattcagatcgaatataatgtagagcagaggacccctcaaaacatgcacctggggcct 1319
 QY 181 cctcagcaatggatggccctggacccctcctcttaagacctcagagctataatc 240
 Db 1320 cctcagcaatggatggccctggatccctcttaagacctcagagctataatc 1379
 QY 241 tttaactcctcttgagcccttcatcctcaactcctcttgaagttgtcctccagat 300
 Db 1380 gttactcctcttgagcccttcatcctcaactcctcttgaagttgtcctccagat 1439
 QY 301 tgaagctgtaagctacaataatgtctcacaatggaaacccagatgagtcacatgaa 360
 Db 1440 tgaagctgtaagctacaataatgtctcacaatggaaacccagatgagtcacatgaa 1499
 QY 361 aatcaccgtgagccctcctggaccggcctcctcagatagctctgtatgttaatgata 420
 Db 1500 gatccaccgtgagccctcctggaccggcctcctcagatagctctgtatgttaatgata 1559
 QY 421 agtaccctcccgaggaaatctcactgcacaaacccctactacactcaattcagtag 480
 Db 1560 agtaccctcccgaggaaatctcactgcacaaacccctactacactcaattcagtag 1619
 QY 481 aagcagttagagcagttgtcagccaactccccaacagtaactgggtttccctgttga 540
 Db 1620 aagcagttagagcagttgtcagccaactccccaacagtaactgggtttccctgttga 1679
 QY 541 ggggtgactgagagcagagctagctggaattcctcagctgactaagatccmaagcct 600
 Db 1680 ggggtgactgagagcagagctagctggaattcctcagctgactaagatccmaagcct 1739
 QY 601 aactggaaggtgacgcgcatcactcttaaaacatgggctgcaactagctcacaccg 660
 Db 1740 aactggaaggtgacgcgcatcactcttaaaacatgggctgcaactagctcacaccg 1799
 QY 661 accaatcagagagctcactaaatgtcactaagcgaacaaacagagtgtaagcaatagcc 720
 Db 1800 accaatcagagagctcactaaatgtcactaagcgaacaaacagagtgtaagcaatagcc 1859
 QY 721 aatcactatgtcctggaagcagcagcggaaggaagatgtggatataactaactgca 780
 Db 1860 aatcactatgtcctggaagcagcagcggaaggaagatgtggatataactaactgca 1919
 QY 781 ttcaagccagcacaacacccctcttggtccctcccatgtatgtggaagctcgtttt 840
 Db 1920 ttcaagccagcacaacacccctcttggtccctcccatgtatgtggaagctcgtttt 1979
 QY 841 cactctattcaactcattaaatcatgcaactgca 875
 Db 1980 cactctattcaactcattaaatcatgcaactgca 2014

RESULT 6
 AAX77526
 ID AAX77526 standard; cDNA: 2946 BP.
 XX
 XX AAX77526;
 AC
 AC 10-AUG-1999 (first entry)
 DT
 DT Human secreted protein AJ172_2 cDNA.
 DE
 DE Secreted protein; testes; brain; blood; placenta; human; murine; thymus;
 KW bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;
 KW cell proliferation; cell differentiation; suppressor; tumour inhibition;
 KW haematopoietic regulator; activin; inhibin; chemotactic; chemokinetic;
 KW haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;
 KW cadherin; tumour invasion suppressor; gene therapy; tissue growth; ss.
 XX
 OS Homo sapiens.
 XX

This is the human secreted protein AJ172-2 nucleotide sequence, obtained from a human adult testes cDNA library. The invention relates to secreted human and murine proteins. The polynucleotides and proteins are predicted to have biological activities which would make them suitable for treating, preventing or ameliorating medical conditions in humans and animals. Detection of the levels of the proteins can be used for the diagnosis of e.g. pre-eclampsia, placental pathology or cancer. Agents which modulate the expression or function of the proteins may be used for treating a neoplastic disease and inhibiting metastasis. Other suggested activities include nutritional activity (e.g. in feeds), cytokine and cell proliferation/differentiation activity, immune stimulating (e.g. as vaccines) or suppressing activity, activity/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, cadherin/tumour invasion suppressor activity, and tumour inhibition activity. The polynucleotide sequences are also stated to be useful for gene therapy.

Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other:

Query Match 55.1%; Score 731.8; DB 21; Length 2946;

Best Local Similarity 90.7%; Pred. No. 2,9e-229; Mismatches 69; Indels 12; Gaps 1;

Matches 794; Conservative 0; Mismatches 69; Indels 12; Gaps 1;

QY 1 tcaaaatcgaagagcttagacttgcttaacgcgcaaaagaggggagacgtttattttt 60
Db 2067 tcaaaatcgaagagcttagacttgcttaacgcgctgaagaggggagacgtttattttt 2126
QY 61 agggagaagaagctggttgatgttaataatcgtgaatcttactgtgaagaagttaaga 120
Db 2127 agggagaagaagctggttgatgttaataatcgtgaatcttactgtgaagaagttaaga 2186
QY 121 aattggaatcgaatataatgtagcagagagacctcaaacactgacactgagggcct 180
Db 2187 aattcgaatcgaatataatgtagcagagagacctcaaacactgacactgagggcct 2246
QY 181 cctcgaacatgtagctgacctggaactcctccttcttagagaccttagagataataatt 240
Db 2247 cctcgaacatgtagctgacctggaactcctccttcttagagaccttagagataataatt 2306
QY 241 tttaactctcttggaacctgatacttcaactctctgttaagtgtgtcttccagat 300
Db 2307 gctactctctcttggaacctgatacttcaactctctgttaagtgtgtcttccagat 2366
QY 301 tgaagctgtaagacacaaatagttctcaaatggaaccagagatgagctcatgactaa 360
Db 2367 cgaaagctgtaaaacta-----caaatggaagccaaagatgagctcatgactaa 2414
QY 361 aatcaccgtgagccctcgagccgagctgtagactatgctctgtaagttaatgacatga 420
Db 2415 gatcaccgagacccctcgagccgagctgtagaccagactatgtaagttaatgacatga 2474
QY 421 agtcacccctcggaggaatctcaactgacacaccccttcaatctcaatcttaagttag 480
Db 2475 agtcacccctcggaggaatctcaactgacacaccccttcaatctcaatcttaagttag 2534
QY 481 aagcagtttagagcagttctcagccaacctcccaacagactgttggtttctctgttga 540
Db 2535 aagcagtttagagcagttctcagccaacctcccaacagactgttggtttctctgttga 2594
QY 541 gggtgagctgagagacagagactgtgattctcctagctgactaagaatcccaagcct 600
Db 2595 tgggtgagctgagagacagagactgtgattctcctagctgactaagaatcccaagcct 2654
QY 601 anctggagaggtgacagatcatctttaaacaagggtgtgacacttagctacacccg 660
Db 2655 agctggagaggtgacagatcatctttaaacaagggtgtgacacttagctacacccg 2714
QY 661 accaatcgaagagctcatcctaataatgctaatacagcaaaacagagagtaagcaataagc 720
Db 2715 accaatcgaagagctcatcctaataatgctaatacagcaaaacagagagtaagcaataagc 2774

QY 721 aatcatctatgctgagagcagcgagggaagacaagatgggataataactcagga 780
Db 2775 aatcatctatgctgagagcagcgagggaagacaagatgggataataactcagga 2834
QY 781 ttcaagcagaagaacagacccctttggtgtccctccatctgatatggagctgtttt 840
Db 2835 ttcaagcagaagaacagacccctttggtgtccctccatctgatatggagctgtttt 2894
QY 841 cactcttcaactatcaatcaatgcaactgca 875
Db 2895 catgctattcaactatcaatcaatgcaactgca 2929

RESULT

AAFS5630
ID AAF55630 standard; DNA: 2781 BP.

AC AAF55630;

DT 29-MAY-2001 (first entry)

DE Nucleotide sequence of a human endogenous retrovirus envelope protein.

KW Envelope protein; HERV; syncytia formation; placental development;

KW syncytia; cancer; cell adhesion; ss.

OS Human endogenous retrovirus.

FH Key location/Qualifiers

FT CDS 762..2378

FT /"tag" a

FT /product= "envelope protein"

PN MO200116171-A1.

PD 08-MAR-2001.

PF 01-SEP-2000; 2000WO-FR02429.

PR 01-SEP-1999; 99FR-0011141.

PR 15-SEP-1999; 99FR-0011793.

PA (INMR) BIO MERIEUX.

PA (INRM) INST NAT SANTE & RECH MEDICALE.

PI Mallet F, Cosset F, Blond J, Lavillette D, Bouton O, Ruggieri A;

DR WPI: 2001-226676/23.

DR P-PSDB: AAB67652.

PT Detecting expression of human endogenous retrovirus envelope protein in

PS cells of a tissue or culture, from its ability to induce syncytia -

XX Disclosure; Page 44-45; 57pp; French.

XX The present sequence encodes a human endogenous retrovirus envelope

CC protein. The specification describes a method for detecting expression

CC of an envelope protein from a human endogenous retrovirus (HERV), in

CC cells, of a tissue or culture. The method comprises detecting syncytia

CC formation due to the fusogenic properties of the envelope protein.

CC Envelope polypeptides and polynucleotides are used to produce

CC therapeutic or prophylactic compositions, particularly for treatment of

CC cancer, to correct defects in placental development (or other natural

CC formation of other types of syncytia), and to promote adhesion of cells

CC in grafts or cellular repair processes. Expression of sequences

CC antisense to the polynucleotide are used to prevent formation of

CC syncytia.

XX

SQ Sequence 2781 BP; 747 A; 770 C; 556 G; 708 T; 0 other:

Query Match 54.9%; Score 730.2; DB 22; Length 2781;

Best Local Similarity 90.6%; Pred. No. 9,5e-229;

Matches	793; Conservative	0; Mismatches	70; Indels	12; Gaps	1;
QY	1	tcaaatcgaaagctttagtactgttaacccgcaaaagaggaggaacccgttattttt	60		
DB	1901	tcaaatcgaaagctttagtactgttaacccgcaaaagaggaggaacccgttattttt	1960		
QY	61	aggaggaagatgctgtttagtattgaatcaatcgtgaatcattacgagaaatlaaga	120		
DB	1961	aggaggaagatgctgtttagtattgaatcaatcgtgaatcattacgagaaatlaaga	2020		
QY	121	aatttagatcgataataatgtatagacagaggaactcaaacactgcaccctgggacct	180		
DB	2021	aatttagatcgataataatgtatagacagaggaactcaaacactgcaccctgggacct	2080		
QY	181	cctcagcaatgatgacctgagactcctccctcttaggaacccctagcagctataatt	240		
DB	2081	cctcagcaatgatgacctgagactcctccctcttaggaacccctagcagctataatt	2140		
QY	241	tttactcctctttagaccctgtatcttcaactcctctgttaagtgtctctccagaat	300		
DB	2141	gtactcctcctttagaccctgtatcttcaactcctctgttaagtgtctctccagaat	2200		
QY	301	tgaagctgtaaagctcaaatagttctcaaatgtgaacccagatgcagtcagactaa	360		
DB	2201	cgaagctgtaaacta-----caaatgtgaacccagatgcagtcagactaa	2248		
QY	361	aattcaccgtgagacccttgaccgctgttagactatgctcgtatgttaatacttga	420		
DB	2249	gattcaccgtgagacccttgaccgctgttagactatgctcgtatgttaatacttga	2308		
QY	421	agtcaccctcccgagaaatcctcaactgcacaaccccttactacatcccaatcagtag	480		
DB	2309	agtcaccctcccgagaaatcctcaactgcacaaccccttactacatcccaatcagtag	2368		
QY	481	aagcagtttagagagttgtcagccaacccctcccaacagacttggttttccctgtgaga	540		
DB	2369	aagcagtttagagagttgtcagccaacccctcccaacagacttggttttccctgtgaga	2428		
QY	541	gggtgactgagagagagactagctgagatttctcctagctgactaagaatccnaagcct	600		
DB	2429	tgaggagactgagagagagactagctgagatttctcctagctgactaagaatccnaagcct	2488		
QY	601	anctgggaaggtgacgcacatcctttaaacaatggggttgcactatgctcaacccg	660		
DB	2489	agctgggaaggtgacgcacatcctttaaacaacggggttgcactatgctcaacccg	2548		
QY	661	accaatcagagagctactaataatgtatcagcaaaaacagagatgaagaataagcc	720		
DB	2549	accaatcagagagctactaataatgtatcagcaaaaacagagatgaagaataagcc	2608		
QY	721	aatcatctattgctctgagagcagcgaggaagagaaagattgggatatataactcagga	780		
DB	2609	aatcatctattgctctgagagcagcgaggaagagaaagattgggatatataactcagga	2668		
QY	781	ttaacagcagcaaacagcaaccccttgggttccctcccaatgtatgagagcctgtttt	840		
DB	2669	ttaacagcagcagcaaacagcaaccccttgggttccctcccaatgtatgagagcctgtttt	2728		
QY	841	cactctatttcaactatataatcaatcagcaactgca	875		
DB	2729	cactctatttcaactatataatccttgcacactgca	2763		

KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	
OS	Homo sapiens.
PN	WO2001/5067-A2.
XX	
PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEO INC.
XX	
PI	Drmannac RT, Liu C, Tang YT;
XX	
DR	WPI; 2001-639362/73.
DR	P-PSDB; ABG20023.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity
XX	
PS	Claim 1; SEQ ID No 20014; 103pp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. AAS64197-AAS94564 represent novel human
CC	diagnostic coding sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published.pct_sequences.
XX	
SO	Sequence 6394 BP; 1840 A; 1597 C; 1384 G; 1571 T; 2 other;
Query Match	54.8%; Score 728.6; DB 23; Length 6394;
Best Local Similarity	78.6%; Pred. No. 5.3e-228;
Matches 1094; Conservative	0; Mismatches 223; Indels 74; Gaps 16;
QY	1 tcaaatcgaaagctttagtactgttaacccgcaaaagaggaggaacccgttattttt
DB	1175 tcaaatcgaaagctttagtactgttaacccgcaaaagaggaggaacccgttattttt
QY	61 aggaggaagatgctgtttagtattgaatcaatcgtgaatcattacgagaaatlaaga
DB	1235 aggaggaagatgctgtttagtattgaatcaatcgtgaatcattacgagaaatlaaga
QY	121 aatttagatcgataataatgtatagacagaggaactcaaacactgcaccctgggacct
DB	1295 aatttagatcgataataatgtatagacagaggaactcaaacactgcaccctgggacct
QY	181 cctcagcaatgatgacctgagactcctccctcttaggaacccctagcagctataatt
DB	1355 cctcagcaatgatgacctgagactcctccctcttaggaacccctagcagctataatt

DB 6900 cctcagccratagatgcccgtgattctccctctcttagagccctctagcagctataatt 6959
QY 241 ttatccctcttggagccctgtatcttcaactccttgttaattgtctctccagaat 300
DB 6960 gctaaccctcttggagccctgtatcttcaactccttgttaattgtctctccagaat 7019
QY 301 tgaagctgaagctaaatagttcttcaatggaaacccagatgcagtcagtaactaa 360
DB 7020 cgaagctgtaacta-----caatggagcccaagatgcagtcagtaactaa 7067
QY 361 aatctacgttggagccctggagccctgttagactatgctctgattgtaacttga 420
DB 7068 gactaacgagagccctggagccctgttagactatgctctgattgtaacttga 7127
QY 421 agtaccctctccgaggaatctcaactgacaaacccctactcaactcaatcagtag 480
DB 7128 aggcaccctctggaggaatctcaactgacaaacccctactcaactcaatcagtag 7187
QY 481 aagcagttagagcagttgtcagcgaacccctcccaacagtaactgggttttcgttga 540
DB 7188 aagcagttagagcagttgtcagcgaacccctcccaacagtaactgggttttcgttga 7247
QY 541 gggtagctggagagagactagctggattctcctagagctgaactgaagaatccnaagc 600
DB 7248 tgggagctggagagagactagctggattctcctagagctgaactgaagaatccnaagc 7307
QY 601 anctggagaggtgacgcgcattctttaaactggggtctgaacttagctacacccg 660
DB 7308 agstggagaggtgacgcgcattctttaaactggggtctgaacttagctacacccg 7367
QY 661 accaactgagagctactactaaatgctaatcagcgaagagagtaagtaagtaagc 720
DB 7368 accaactgagagctactactaaatgctaatcagcgaagagagtaagtaagtaagc 7427
QY 721 aatcatctattgctctgagagacagcgaggagacaagattggatataactcagga 780
DB 7428 aatcatctattgctctgagagacagcgaggagacaagattggatataactcagga 7487
QY 781 ttcaagcagagcaaacaccccttgggtccctccctctgtaggggctgtttt 840
DB 7488 ttcaagcagagcaaacaccccttgggtccctccctctgtaggggctgtttt 7547
QY 841 cactctattcactctatataatcagtaactgca 875
DB 7548 catgctattcactctatataatcagtaactgca 7582
SULF 11
ID AA59215 standard; DNA; 7582 BP.
AC AA59215;
DT 07-NOV-2000 (first entry)
XX Human endogenous retrovirus W (HERV-W) sequence.
DE Human endogenous retrovirus W (HERV-W) sequence.
XX Autoimmune disease: retrovirus; human endogenous retrovirus W; HERV-W;
KW gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
XX Human endogenous retrovirus.
OS
XX
FH Key Location/Qualifiers
FT LTR 1..120
FT LTR /tag= a
FT LTR /note= "R of 5' LTR"
FT LTR 121..575
FT LTR /tag= b
FT LTR /note= "U5 of 5' LTR"
FT primer_bind 579..596
FT CDS /tag= c
FT CDS 5581..7194

FT /tag= d
FT /note= "ORF1 env538"
FT CDS 7039..7194
FT /tag= e
FT /note= "ORF2 52 AA"
FT CDS 7112..7255
FT /tag= f
FT /note= "ORF3 48 AA"
FT misc_feature 7244..7254
FT /tag= g
FT /note= "polypurine tract"
FT LTR 7256..7582
FT /tag= h
FT /note= "U3-R of 3' LTR"
FT polyA_signal 7563..7569
FT /tag= i
PN W0200043521-A2.
XX
PD 27-JUL-2000.
XX
PF 21-JAN-2000; 2000MO-FR00144.
XX
PR 21-JAN-1999; 99PR-0000888.
XX
PA (INMR) BIO MERIEUX.
PI Paranhos-Baccala G, Mallet F, Voisset C;
XX
XX WPI: 2000-499229/44.
XX
PT New nucleic acid from human endogenous retrovirus, useful e.g. for
PT diagnosis of autoimmune disease and complications of pregnancy,
PT contains at least part of the gag gene
XX
XX
PS Disclosure; Page 49-52; 53pp; French.
XX
XX
CC The present sequence represents an endogenous retrovirus, which is
CC associated with an autoimmune disease, and is integrated into the human
CC genome. The retrovirus is human endogenous retrovirus W (HERV-W). The
CC HERV-W retrovirus is associated with autoimmune disease, failure of
CC pregnancy or disorders of pregnancy. HERV-W nucleic acid fragments, or
CC proteins derived from it, are useful for diagnosis of autoimmune
CC disease (specifically multiple sclerosis) and for monitoring pregnancy.
CC The nucleic acid fragments may also be used for in situ labelling of
CC isolated chromosomes, while the transcription product can be used to
CC study or monitor T cell proliferation in vitro.
XX
SQ Sequence 7582 BP; 2156 A; 1876 C; 1538 G; 1796 T; 216 other;
Query Match 54.4%; Score 722.6; DB 21; Length 7582;
Best Local Similarity 88.5%; Pred. No. 5.5e-726;
Matches 774; Conservative 19; Mismatches 70; Indels 12; Gaps 1;
QY 1 tcaaatcgaaagccttagactgtcgaaccgcaaaagaggggaacctgtatttt 60
DB 6720 tcaaatcgaaagccttagactgtcgaaccgcaaaagaggggaacctgtatttt 6779
QY 61 aggggaagatgctgttagatgttaatcaatctggaatcattcagtagaagttaaga 120
DB 6780 aggggaagatgctgttagatgttaatcaatctggaatcattcagtagaagttaaga 6839
QY 121 aattgagatcgaatataatgttagagagagacctaacaacacgacccctgggct 180
DB 6840 aattgagatcgaatataatgttagagagagacctaacaacacgacccctgggct 6899
QY 181 cctcagcgaatgagtcgctggagactccctctttagagacccctagcagctataatt 240
DB 6900 cctcagcgaatgagtcgctggagactccctctttagagacccctagcagctataatt 6959
QY 241 ttactcctcttggagccctgtatctcaactccttgttaattgtctctccagaat 300

D6 6960 gctactcctcttggaccctgtatctttracccctcttacttctgtcttccagaat 7019
QY 301 tgaagctgtaaagctacaataatgtcttctcaaatggaaccagatcgatccatgactaa 360
D6 7020 cgaagctgtiraacta-----caaatggaagcccaagaatgctgcgtccaagactaa 7067
QY 361 aatcaccctgtggaacccctgtgacccgctgtctagactatgtctgtatgtatgacatgta 420
D6 7068 gatcaccagcagaccctgtgacccgctgtctagcaccagactgtgtatgtatgacatcaa 7127
QY 421 agtcccccctccgaggaatctctacacacccctactactactactactactactactactg 480
D6 7128 aggcacccctccctcgaggaatctctacacacccctactactactactactactactactg 7187
QY 481 aagcagtttagagcaggtgtctcagcaccactcccaacagactactgtgttctctgttga 540
D6 7188 aagcagtttagagcaggtgtctcagcaccactcccaacagactactgtgttctctgttga 7247
QY 541 ggtgtgactgagagacagagactgtgattctctagctgactaagaatcccaagcct 600
D6 7248 tggggagactgagagacagagactgtgattctctagctgactaagaatcccytaagcct 7307
QY 601 aacttggaaggttgaccgataccatctttaaacaatgagggctgtgcaactagctacaacgg 660
D6 7308 aggtggagaggttgaccgataccatctttaaacaacgggctgtgcaactagctacaacgg 7367
QY 661 accaatcagagagctcactataaattgtctaatacagcaaaacagagaggttaagcaatagcc 720
D6 7368 accaatcagagagctcactataaattgtctaatacagcaaaacagagaggttaagcaatagcc 7427
QY 721 aatcatctatgtcctgtgagagacagcgaggaaggaatgttgatataaaccagga 780
D6 7428 aatcctatctgtcmgtgagagacagcgaggaaggaatgttgatataaaccagga 7487
QY 781 ttcaagcagacagacagacaccccttgggtccctccatgtgtatgtgagactgtttt 840
D6 7488 ttgcagcggcagacagacaccccttgggtccctccatgtgtatgtgagactgtttt 7547
QY 841 cactctatctcactctataaatacagcaactgca 875
D6 7548 catgtattctcactctataaatacagcaactgca 7582
RESULT 12
ID AAX25660 standard; cDNA to mRNA; 1136 BP.
AC AAX25660;
21-MAY-1999 (first entry)
XX Human endogenous retrovirus W clone cl.C4C5.
DE Human endogenous retrovirus W clone cl.C4C5.
XX Clone; human endogenous retrovirus; genome; autoimmune disease;
KW multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
KW disseminated lupus erythematosus; pregnancy; chromosomal marker: ss.
OS Human endogenous retrovirus.
XX OS
XX WO9902696-A1.
PN 21-JAN-1999.
XX 21-JAN-1999.
XX 06-JUL-1998; 98WO-FR01442.
XX 07-JUL-1997; 97FR-0008815.
XX (IMMR) BIO MERIEUX.
XX Beeseme F, Blond JL, Bouton O, Mallet F, Mandrand B;
XX WPI; 1999-120897/10.
XX

PT New nucleic acid sequences from human endogenous retrovirus W -
PT expressed exclusively in placenta and useful in diagnosis and
PT therapy of autoimmune disease, and abnormal or failed pregnancy
XX
XX
PS Claim 1; Page 59-60; 106pp; French.
XX
XX This sequence represents clone cl.C4C5 of the human endogenous retrovirus
CC (HERV) W genome. The nucleic acids, their fragments or peptides encoded
CC by them are markers of autoimmune disease (e.g. multiple sclerosis,
CC rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-
CC dependent diabetes and related pathologies) and of abnormal or
CC unsuccessful pregnancy and can be used as chromosomal markers for
CC susceptibility to these conditions, or proximity markers of genes
CC associated with this susceptibility.
XX
SO Sequence 1136 BP; 336 A; 289 C; 241 G; 270 T; 0 other;
Query Match 54.38; Score 721.2; DB 20; Length 1136;
Best Local Similarity 90.08; Pred. No. 5.1e-226;
Matches 787; Conservative 0; Mismatches 75; Indels 12; Gaps 1;
QY 1 tcaaatcgaagagcttagactgtctaacccgcaaaagagggagactgttatttt 60
D6 254 tcaaatcgaagagcttagactgtctaacccgcaaaagagggagactgttatttt 313
QY 61 aggggaagaagctgtgtatgtatcaatctggaatctctgaaagctgaaagta 120
D6 314 aggggaagaagctgtgtatgtatcaatctggaatctctgaaagctgaaagta 373
QY 121 aatttgaatctgaaatataatgtatgagacagagacacttcaaaacactgacccctgggct 180
D6 374 aattcgaatctgaaatataatgtatgagacagagacacttcaaaacactgacccctgggct 433
QY 181 cctcagcaaatggatccctctgacccctccctctttagagactttagagactataat 240
D6 434 cctcagcaaatggatccctctgacccctccctctttagagactttagagactataat 493
QY 241 ttactcctcttggaccctgtatcttcaactctctgttgaattgtctctccagat 300
D6 494 gctactccttctggaccctgtatcttcaactctctgttgaattgtctctccagat 553
QY 301 tgaagctgtaaagctacaataatgtcttcaaatggaacccagatgtccatgactaa 360
D6 554 cgaagctgtaaacta-----caaatggaagcccaagatgacgtccaagactaa 601
QY 361 aatcaccgttgaccctgtgacccgctgtcagactaagctgtatgtatgactgca 420
D6 602 gatctaccgacagaccctgtgacccgctgtcagactaagctgtatgtatgactgca 661
QY 421 agtaccctcccgaggaatctcaactgacaaacccctactactactactactactag 480
D6 662 agtaccctccctctgaggaatctcagctgcacacccctactactactactactag 721
QY 481 aagcagtttagagcaggtgtctgacacacccctcccaacgactgtgtgttctctgtga 540
D6 722 aagcagtttagagcaggtgtctgacacacccctcccaacgactgtgtgttctctgtga 781
QY 541 ggtgtgactgagagacagagactgtgattctctagctgactaagaatcccaagcct 600
D6 782 tggggagactgagagacagagactgtgattctctagctgactaagaatcccaagcct 841
QY 601 aacttggaaggttgaccgataccatctttaaacaatgagggctgtgcaactagctacaacgg 660
D6 842 agcttggaaggttgaccgataccatctttaaacaacgggctgtgcaactagctacaacgg 901
QY 661 accaatcagagagctcactataaattgtctaatacagcaaaacagaggttaagcaatagcc 720
D6 902 accaatcagagagctcactataaattgtctaatacagcaaaacagaggttaagcaatagcc 961
QY 721 aatcatctatgtcctgtgagacagcgaggaaggaatgttgatataaaccagga 780
D6 962 aatcatctatgtcagagacagcgaggaaggaatgttgatataaaccaggtc 1021

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 20, 2002, 09:52:31 ; Search time 4429.18 Seconds
(Without alignments)
6279.123 Million cell updates/sec

Title: US-09-319-156A-12
Perfect score: 1329
Sequence: 1 tcaaatcgagagacttag.....gcaagatccccagtaaca 1329

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenBankl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_cm:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_inv:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
------------	-------------	--------	-------	-------------

1	1326	99.8	1329	6	AX001030	AX001030 Sequence
2	1316.4	99.1	163803	9	AC093531	AC093531 Homo sapi
3	1035.2	77.9	56093	6	AX329572	AX329572 Sequence
4	1035.2	77.9	56093	9	HSAC000064	AC000064 Human BAC
5	1035.2	77.9	149194	9	AC007566	AC007566 Homo sapi
6	1023.2	77.0	10499	6	AX007980	AX007980 Sequence
7	1006.6	75.7	188919	2	AC009727	AC009727 Homo sapi
8	989.6	74.5	161571	2	AC092708	AC092708 Homo sapi
9	989.6	74.5	166700	9	AC013759	AC013759 Homo sapi
10	989.6	74.5	186911	2	AP001488	AP001488 Homo sapi
11	985.6	74.2	118241	9	AP001599	AP001599 Homo sapi
12	985.6	74.2	340000	9	AP001697	AP001697 Homo sapi
13	984	74.0	91097	9	AP001600	AP001600 Homo sapi
14	982.6	73.9	172721	2	AC008121	AC008121 Homo sapi
15	979.2	73.7	145785	2	AC016173	AC016173 Homo sapi
16	968.8	72.9	157711	9	AC055738	AC055738 Homo sapi
17	968.8	72.9	207945	9	CNS01D5S	AL121841 Human chr
18	951.4	71.6	169072	9	AC009946	AC009946 Homo sapi
19	951.2	71.6	162522	9	AC093839	AC093839 Homo sapi
20	948.2	71.3	155925	9	AC096536	AC096536 Homo sapi
21	912.8	68.7	183680	9	AC098859	AC098859 Homo sapi
22	912.8	68.7	221525	2	AC021652	AC021652 Homo sapi
23	876.4	65.9	174019	9	AP001538	AP001538 Homo sapi
24	876.4	65.9	340000	9	AP001674	AP001674 Homo sapi
25	870.4	65.5	83412	9	AC092843	AC092843 Homo sapi
26	859	64.6	169462	2	AC024033	AC024033 Homo sapi
27	851.2	64.0	251124	9	HUAE000660	AB000660 Homo sapi
28	833.8	62.7	140756	9	AL139038	AL139038 Human DNA
29	833.8	62.7	183499	2	AL607153	AL607153 Homo sapi
30	833	62.7	839	9	AF127226	AF127226 Homo sapi
31	824.8	62.1	114455	9	AC093742	AC093742 Homo sapi
32	816.6	61.4	187399	9	AC067942	AC067942 Homo sapi
33	811.4	61.1	153104	2	AC008564	AC008564 Homo sapi
34	809.8	60.9	114972	9	AC016617	AC016617 Homo sapi
35	808.2	60.8	181016	2	AC073129	AC073129 Homo sapi
36	801.6	60.3	156821	9	AC008448	AC008448 Homo sapi
37	798.2	60.1	88328	9	AL357874	AL357874 Human DNA
38	789.6	59.4	172182	2	AC017114	AC017114 Homo sapi
39	777.6	58.5	188038	2	AC092510	AC092510 Papio cyn
40	768	57.8	164310	2	AC016156	AC016156 Homo sapi
41	768	57.8	182224	9	AC093116	AC093116 Homo sapi
42	766	57.6	117124	2	AC110614	AC110614 Homo sapi
43	762.4	57.4	123253	9	AC078963	AC078963 Homo sapi
44	762.4	57.4	138003	2	AC018346	AC018346 Homo sapi
45	762.4	57.4	158059	2	AC079947	AC079947 Homo sapi

ALIGNMENTS

RESULT	1	1329 bp	DNA	linear	PAT 10-MAR-2000
AX001030	AX001030	Sequence 12 from Patent WO9902666.			
LOCUS	AX001030				
DEFINITION	Sequence 12 from Patent WO9902666.				
ACCESSION	AX001030				
VERSION	AX001030.1	GI:7241266			
KEYWORDS					
SOURCE					
ORGANISM	unidentified.				
	unclassified.				
REFERENCE	1 (bases 1 to 1329)				
AUTHORS	Ott,C. and Bedin,F.				
TITLE	RETROVIRAL NUCLEIC MATERIAL AND NUCLEOTIDE FRAGMENTS, IN PARTICULAR ASSOCIATED WITH MULTIPLE SCLEROSIS AND/OR RHEUMATOID ARTHRITIS, FOR DIAGNOSTIC, PROPHYLACTIC AND THERAPEUTIC USES				
JOURNAL	PATENT: WO 9902666-A 12 21-JAN-1999.				
FEATURES	BIO MERIEUX (FR); OTT CATHERINE (FR)				
	Location/Qualifiers				
	1..1329				
	/organism="unidentified"				
	/db_xref="taxon:32644"				

BASE COUNT	346 a	352 c	288 g	340 t	3 others
ORIGIN					

Query Match 99.8%; Score 1326; DB 6; Length 1329;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1329; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 tcaaatcgaagagcttagacttgctaaaccgcaaaagaggaggaacgcttattt 60
    |||||||
DB 1 TCAAAATCGAAGAGCTTAGACTTGCTAACCGCCAAAAGAGGGGAACTCTTTATTTT 60
QY 61 aggggaagaagctgcttagatgttattcaatctcgaaatcactgagaaagttaaga 120
    |||||||
DB 61 AGGGGAAGAAAGCTGTATGTATGTAATCAATCGAATCTTACTGGAATTAAGA 120
QY 121 aatttgagatcgaaatataatgttagagcagagagacctcaaaacactgcaacctgggacct 180
    |||||||
DB 121 AATTGGAATCGAATATATATATAGAGCAGAGAGACCTTCAAAACACTGCACTGGGGCT 180
QY 181 cctcaagcaatgagatgctcctgagactcctcctctttagagcctttagagactataatatt 240
    |||||||
DB 181 CCTCAGCCAAATGATGCTGCTGAGACTCTCCCTCTTAGAGACTCTAGCAGCTATATATAT 240
QY 241 tttaactcctcttgagacccttactctcaactctcttgcttgaagtgtctcttcagagat 300
    |||||||
DB 241 TTTACTCTCTTTGGACCCCTTATCTTCACTTCTTGTAAAGTTTGTCTTCCAGAT 300
QY 301 tgaagctgttaagctacaaatagttcttcaaaatggaacccagatgacatcactgactaa 360
    |||||||
DB 301 TGAAGCTGTAAAGCTACAAATAGTTCTTCAATGGAACCCAGATGACATGACTTAA 360
QY 361 aatctaccgttgagaccccttgagacggccctgctgaagtactgctgtaagttaatgacattga 420
    |||||||
DB 361 AATCTACCGTGGACCCCTGAGACGGCTGCTAGACTATGCTGATGTAAATGACATTGA 420
QY 421 agtcaacccctccgaggaatctcaactgcaaacccctactaactcctcaatctcaatga 480
    |||||||
DB 421 AGTCACCCCTCCGAGGAAATCTCACTGCAACACCCTTATACACTTCATTCAGTAG 480
QY 481 aagcagttagagcagttgttcagcgaacctcccaacagttacttgggtttctcgttga 540
    |||||||
DB 481 AAGCAGTAGAGCAGATTGTCTAGCCAACTCCCAACAGTACTTGGGTTTCTGTTGAGA 540
QY 541 ggggtgagctgagagagagactagctggaatttcttagctgactaagaatcccaagact 600
    |||||||
DB 541 GGGGTGAGCTGAGAGACAGACTAGCTGATTTCTTAGGCTGACTAAGATCCCAAGCT 600
QY 601 anctgggagagtgagacgcatcatttaaacatgagggctgtgcaacttagctcacaacgg 660
    |||||||
DB 601 ANCTGGGAAGGTGACCGCATCTTTAAACATGGGGCTTGCAACTTAGCTCACACCG 660
QY 661 accaatcagagagctcactaaatgctaaatcaggaacaaagaggtaaaagcaatagcc 720
    |||||||
DB 661 ACCAATCAGAGAGCTCAATAAATGCTAATCAGGCMAAAMACAGAGGTAAGCAATAGCC 720
QY 721 aatatcatctgctctgagagacagcgagggaaggaagatctgggataataacacagaca 780
    |||||||
DB 721 AATATCTATTGCTCTGAGAGACAGCGGGAAGACAAAGATGGGATTTAACTCAGGCA 780
QY 781 ttcaagccaagacaagacaccccttttgggtccctcccatctgtatgagagctgtttt 840
    |||||||
DB 781 TTCAAGCCAGACAGACAGAACCCCTTTGGGTGCCCTCCCATTTGATGGAGACTCTGTTT 840
QY 841 caactcatctcactcattaaatcagcaactgcaactcttctgtctcgttctttatgg 900
    |||||||
DB 841 CACTCTATTACTCTATTAAATCATGCACTCTCTGTGCTCCCTGTTTATAGG 900
QY 901 ctcaagctgagcttttctgacatcccaactgctgttgcacacgctcacaagccgact 960
    |||||||
DB 901 CTCAGGCTGACTTTTGTGCGCATCCACACTGCTGTTTCCACCGCAGACAGCCGCT 960
QY 961 gctgactcctcactccttggatcagcagagtgctcaactgctgctcctgatacagaagct 1020
    |||||||
DB 961 GCTGACTCTCATCCCTTTGGATCCAGAGAGTGTGCATGTGCTGATCCTGATCCAGAGAGT 1020
  
```

```

QY 1021 accactgcccactcccgatcagcgttaagagcttgccattgtctcgtcagtgtaagtgcc 1080
    |||||||
DB 1021 ACCATTGCCACTCCCGATCAGGCTTAAGGCTTGCAATGTCTGATGCTGAAGTGC 1080
QY 1081 tgggttgccttaataagaaactgcaactgttcactgttggttcatagttctcttcatagac 1140
    |||||||
DB 1081 TGGGTTTGTCTTAATATAGAACTGAACTGCTGCTGCTTCCATGATGAC 1140
QY 1141 ccagggctcttaataagagactaactcactcagcagatgcccagaattccattccttgta 1200
    |||||||
DB 1141 CCAGGGCTTCTTAATAGACTTAACTACCTCACCAGATGCCCCAAGATTCATTCTTGTA 1200
QY 1201 tctgtgagggccaagaaccccgagtcagagaagtgtgagcttgcacatttgggaagtgg 1260
    |||||||
DB 1201 TCTGTGAGGGCCAAGAACCCCGTCAAGAAAGTAGAGCTTGCCACATTGGGAAGTGG 1260
QY 1261 ccaactgcatttctgtgtagcgcccaacacacttctggagctgtggaggaagaatccc 1320
    |||||||
DB 1261 CCCACTGCATTTTGTATAGCGGCCACCACCATCTTGGAGCTGTGGAGCAAGATCCC 1320
QY 1321 ccagtaaca 1329
    |||||||
DB 1321 CCACTAACA 1329
  
```

RESULT 2
 AC093531/c 163803 bp DNA linear PRI 16-NOV-2001
 LOCUS Homo sapiens chromosome 5 clone RP11-405L7, complete sequence.
 DEFINITION AC093531
 ACCESSION AC093531.2 GI:16945981
 VERSION
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 163803)
 DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 163803)
 DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 3 (bases 1 to 163803)
 DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (16-NOV-2001) DOE Joint Genome Institute, 2800 Mitchell
 Drive, Walnut Creek, CA 94598, USA
 COMMENT On Nov 16, 2001 this sequence version replaced gi:15383820.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center
 www.shgc.stanford.edu
 Quality: Phrap Quality >=40 99.8% of Sequence;
 Estimated Total Number of Errors is 0.3.

FEATURES
 source
 1..163803
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="RP11-405L7"
 BASE COUNT 54797 a 31372 c 29086 g 48548 t
 ORIGIN

Query Match 99.1%; Score 1316.4; DB 9; Length 163803;
 Best Local Similarity 99.3%; Pred. No. 0;
 Matches 1320; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```

QY 1 tcaaatcgaagagcttagacttgctaaaccgcaaaagaggaggaacgcttattt 60
  
```

Dy	117604	TTAAATTTGAAGACCTTTAGACTTGTCTAAACGCCAAAAGAGGGGAACTGTTATATTTT	117545
Dy	61	aggggaagaatgctggttagtatgtttaatcaatctggaataactaactgagaagttaaaga	120
Dy	117544	AGGGGAAGAAGTGGTTAGTATGTTAAATCAATCTGGAAATATTACAGAAAGTTAAAGA	117485
Oy	121	aattggaatcgaaataaagttagagagaaggagacctcaaaacactcaccctggggcct	180
Dy	117484	AATTGGATATGAAATATATATATAGAGCAGAGAGACCTTCAAAACACTGACCCCTGGGGCT	117425
Oy	181	cctcaagccaatylgaatgcccctgagactctccctctcttagaactctagcaagctaatat	240
Dy	117424	CCCTAGCCAAATGAGATGGCCCTGAGACTCTCCCTCTTGTAGACCTTACACACTTAAATAT	117365
Oy	241	tttactcctcttggaccctgtatcttcaactctctctgttaagttgtctctccgaat	300
Dy	117364	TTTACTCCTCTTGTGAGACCTGTATCTTTAACTTCCTGTTTAAAGTTGTCTCTCCAGAAAT	117305
Oy	301	tgaaactgtlaaagctacaaatagttcttccaatylgaaccccagaatgacatgacataa	360
Dy	117304	TGAAGCTGTAAAGCTACAAATAGTCTTTCAATATGGAACCCCAATGCAATGCAATGACTAA	117245
Oy	361	aatctacccgtgagcccttgagccggctcgtctagactatgctctgtaagttaatgaactga	420
Dy	117244	AATCTACGCTGAGACCCCTGGACCCGGCTCGTAGAGACTGTGCTGATGTTAATGACATTTGA	117185
Oy	421	agtaacccctccgagaagaatctcaactctgcaaacccctactaactacatccaatgaagg	480
Dy	117184	AGTCAACCCCTCCAGAGAAATCTAACTGCACAAACCCCTACTACACTCCAAITTCAGTAGG	117125
Oy	481	aagcaagttagagcagttgttcagcccaacctcccacaagtaacttggtttctctgttga	540
Dy	117124	AAGCAGTTAGAGCAAGTTGTACGCCAACCTCCCAACAGTACTTGGGTTTCTCTGTAGAGA	117065
Oy	541	gggtggaacttagaagaacagagactatgatttctcttagctgagactaagaatcccaagcct	600
Dy	117064	GGGTGACTGTAGAGACAGACTACTGTGATTTCTTAGGCTGACTAAGAAATCCCTTAACCT	117005
Oy	601	ancttggagaagtgaacccatccactctttaaacaatggggcttgcacacttagctcaacccg	660
Dy	117004	AGCTGGGAAGGTGAACCCGATCCATCTTTTAAACATGGGGGCTTGCAACTTAGTGCACACCG	116945
Oy	661	accaatcagagaagctcaactaaatgtctaatatagggcaaaaacaggaagttaaagcaatagcc	720
Dy	116944	ACCAATCTGAGAAGCTCACTAAAAATGCTAATAGGCAAAAACAGGAGTTAAGCAATATGGCC	116885
Oy	721	aatcatctatgctctggaagacagccgggaagagaacaaggaattgggaatataaactcaagga	780
Dy	116884	AATCATCTATTTGCTGTAGAGACACAGCCGGAAGAGCAAGATGGGATATATAACCCAGGCA	116825
Oy	781	ttcaagccagaacaagaaccccccttgggtccctccccaattgtaaggagcttgttt	840
Dy	116824	TTTCAAGCCAGCAACAGCAACCCCTTTGGGCTCCCTCCCATTTGATGGAGCTCTGTTTT	116765
Oy	841	caactatcttaactctataaataatgcaactgcaactctctgtctcgtttttttatgg	900
Dy	116764	CACCTATTTTCACTCTATTAATCTTTGCACTGTGACCTCTTGATCCGTGTTTTTAATGG	116705
Oy	901	ctcaagctgagacttttgttgcacatccacaacactgtctttagccacccgtcaacagaccgct	960
Dy	116704	CTCAAGCTGAGACTTTTGTGTGCCATTCACACCACTGCTGTGTTGCCACCGTCACAGACCGCT	116645
Oy	961	gctgaactcacaatcccttttggatccaagcaagagtgtccaactgtgtcctctgataccaagcagct	1020
Dy	116644	GCTGACTTTCATCCCTTTGGATCCAGCAAGATGTGCCATGTGGCTCTCGATATCCACGAGGT	116585
Oy	1021	accacattgcacatcccgatacaggtctaaagcctgtccatgtctcctgatatgctcaagcgc	1080
Dy	116584	ACCCATTGCGCACTCCCGATCAAGGCTTAAAGCTTGGCATTGTCTGTGATGGCTTAAGTGC	116525
Oy	1081	tggtgttgtccbaatagaactgaaacactgtgctacatgggttccatlggttctcttccatgac	1140

Db	116524	TGGGTTTGTCCTAATPAGAACTGAACACTGGACACTGGTTCATGTTTCCTTCATGAC	1164655
Qy	1141	ccaaagcgtcttctaataagagcctataacatcaccgcagatgcccagaattccattcccttgta	1200
Db	116464	CCACGGGCTTCTAAATPAGAGCTATTAACACTCACCCGATGGCCCAAGATTCCATTTCTTGTA	1164050
Qy	1201	tctgtgaagccaagaaccccagagtcagagaaagtgaagcttgcaccatttgggaagtgg	1260
Db	116404	TCTGTGAGGCCAATAACCCCAAGCTCAAGAAACGTGAGGCTTGGCACCATTTTGGGAACCTGG	1163455
Qy	1261	ccccatgcacatttggtagcggcccaccaccatcttgggaagcctgttgggaacaaagatccc	1320
Db	116344	CCCCCTGCCATTTTGGTAGAGGCCCCACACCATCTTGGGAGCTCTGGAGCAAGAGATTCCC	1162855
Qy	1321	ccagtaaca	1329
Db	116284	CCAGTAACA	116276

RESULT		3				
LOCUS	AXJ29572		56093 bp	DNA	linear	PAT 09-JAN-2002
DEFINITION	Sequence 81 from Patent WO0194629.					
ACCESSION	AXJ29572					
VERSION	AXJ29572.1		GI:18102550			
KEYWORDS						
SOURCE						
ORGANISM	human.					
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 					
AUTHORS	Young,P.E., Augustus,M., Carter,K.C., Emner,R., Endress,G., Horligan,S., Soppel,D.R. and Weaver,Z. Cancer gene determination and therapeutic screening using signature gene sets					
TITLE	Patent: WO 0194629-A 81 13-Dec-2001;					
JOURNAL	Avalon Pharmaceuticals (US)					
FEATURES						
Source	Location/Qualifiers 1..56093 /oranism="Homo sapientis" /db_xref="taxon:9606"					
BASE COUNT	16164 a 12346 c 10702 g 16881 t					
ORIGIN						
Query Match	77.9%; Score 1035.2; DB 6; Length 56093;					
Best Local Similarity	90.1%; Pred. No.6.le=304;					
Matches 1133; Conservative	0; Mismatches 111; Indels 13; Gaps 2;					
OY 1	tcaaaatcggaagactttagacctgctaaccgccaaaaggagggaacctgttatatttt 60					
Db 37018	TCAAAATCGAAGACTTTAGACTTGCTAACGCGTAAGAAGGGGGGAACCTGTATTATTTT 37077					
OY 61	aggggagaagatgcgttgttgatgatgttaaatcatcctggatatcatacttgaaaaagttaaaga 120					
Db 37078	AGGGGAGAGATGCCTGTTTATTAGTTATAACAATCCGGAATCGTCACTGAAAAAGTTAAAAMA 37137					
OY 121	aattggatatgaatfatatatgtfagagcagaggacctccaanaactgcagccctgggggacct 180					
Db 37138	AATTGAGATGCAATPACACGTAAGACAGAGAGCTTCGAANAACGTGAGACCCTGGGGGCTT 37197					
OY 181	ccttcagacaalbgagcccctggagacctcccccttctttagaacctctlagcacgtataataatt 240					
Db 37198	CCTCAGCCAATBGAGGCCCTGGAATTCGCCCTTTCTTAGAGACCTCTAGCACGTAATATATT 37257					
OY 241	ttactctcctlttggaccctgtatcttcaaacttcccttgyltaagtltyctcltccagaagt 300					
Db 37258	GCTACTCCTCTTTGGACCCCTGTATCTTTMAACTCTCTTGTTAACTTWTGTCCTCCAGAAT 37317					
OY 301	tgaagctgltaaagctacaataatgcttcttcaaatlgaaaccccgagatgcagctcatgactaa 360					
Db 37318	CGAAACCTGTAAAACTA-----CAAATGGAGCCCCAAGATGACATCCCAAGACTAA 37365					

Accession	Species	Gene	Protein	Length	Score	E-value	Identity	Positives	Gaps	Conserved Domains
QY 361	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	420	100.0	0.0	100%	420	0	LOC101928441
Db 37366	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	420	100.0	0.0	100%	420	0	LOC101928441
QY 421	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	480	100.0	0.0	100%	480	0	LOC101928441
Db 37426	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	480	100.0	0.0	100%	480	0	LOC101928441
QY 481	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	540	100.0	0.0	100%	540	0	LOC101928441
Db 37486	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	540	100.0	0.0	100%	540	0	LOC101928441
QY 541	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	600	100.0	0.0	100%	600	0	LOC101928441
Db 37546	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	600	100.0	0.0	100%	600	0	LOC101928441
QY 601	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	660	100.0	0.0	100%	660	0	LOC101928441
Db 37606	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	660	100.0	0.0	100%	660	0	LOC101928441
QY 661	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	720	100.0	0.0	100%	720	0	LOC101928441
Db 37666	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	720	100.0	0.0	100%	720	0	LOC101928441
QY 721	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	780	100.0	0.0	100%	780	0	LOC101928441
Db 37726	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	780	100.0	0.0	100%	780	0	LOC101928441
QY 781	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	840	100.0	0.0	100%	840	0	LOC101928441
Db 37786	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	840	100.0	0.0	100%	840	0	LOC101928441
QY 841	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	900	100.0	0.0	100%	900	0	LOC101928441
Db 37846	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	900	100.0	0.0	100%	900	0	LOC101928441
QY 901	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	960	100.0	0.0	100%	960	0	LOC101928441
Db 37906	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	960	100.0	0.0	100%	960	0	LOC101928441
QY 961	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	1020	100.0	0.0	100%	1020	0	LOC101928441
Db 37966	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	1020	100.0	0.0	100%	1020	0	LOC101928441
QY 1021	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	1080	100.0	0.0	100%	1080	0	LOC101928441
Db 38026	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	1080	100.0	0.0	100%	1080	0	LOC101928441
QY 1081	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	1140	100.0	0.0	100%	1140	0	LOC101928441
Db 38086	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	1140	100.0	0.0	100%	1140	0	LOC101928441
QY 1141	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	1200	100.0	0.0	100%	1200	0	LOC101928441
Db 38146	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	1200	100.0	0.0	100%	1200	0	LOC101928441
QY 1201	<i>Macaca mulatta</i>	LOC101928441	LOC101928441	1257	100.0	0.0	100%	1257	0	LOC101928441
Db 38205	<i>Macaca mulatta</i>	LOC101928								

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Pauley, A.
The sequence of H. sapiens BAC clone RG083M05
unpublished (1996)
2 (bases 1 to 56093)

Department of Genetics, Washington University
St. Louis, MO 63108, USA
e-mail: saplens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

SOURCE INFORMATION:
This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. For references see: Shizuya et al., *Proc. Natl. Acad. Sci.* 93:8794-8797 (1992); Kim et al., *Genomics* 34:213-218 (1996).

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The orientation of this clone is unknown. Actual start of this clone is at base position 1 of H_RG083M05; actual end is at 560933 of H_RG083M05

This clone contains STS SWSS1725

FEATURES

source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-7q22"
/clone="H_RG083M05"
/clone_11p="C17B-9786K-B"
complement(838..1131)
repeat_region

```

CDS

/gene="WUSC:H.RG083M05.1"
 join(<1360..1503,4181..4370,4587..4774,6422..6556,
 9483..9547,11631..11773,11864..12021,1331..13296,
 14885..14988,16349..16546,16837..16971)
 /gene="WUSC:H.RG083M05.1"
 /note="ATPase: strong similarity to peroxisome
 biosynthesis protein PSL1 (P1D:q1172019); coded for by
 human CDNA C04279 (NID:q1467530)"

```
/db_xref="GI:1669371"
```

/translation-"KRENIQKTL

SPDAEQRC EILCNV IKNKLD CDIN

SRQSI STREKLVTTLDFOKALRG

TIQLPAKVCLKKEKYPELEFANLPI

KGBELSKY IGASEQAVRDIE IKA
I I MOI DOVCEI OCYKVI A MENDD

LLI'QLDGV EGLQGVYVLAATSKRPD
OMI HSEI VEP I ETI NVI ENDI PI A

QMLNHFVSKLEILNVLSDSLPLLA
MIJ.SKMSEFI.PDESKENMYBI.YEG

MLLSKMSLLEPDESKFNMIKLEFG
GKDOI.E.SOPPYI.BTASOEGCOET.T

OKDQDLF 3QFF V LK LK3QLEOCQLELI
KTBRLAISOSHI,MTAIGHTBPSITSE

NIKEAIDONIAHONIKU
comment (4948, 5130)

```
/rpt family="ALU"
```

```
complement(6581, .7133)
```

1. The first part of the document is a list of names and titles, including "The Hon. Mr. Justice" and "The Hon. Mr. Justice".


```

repeat_region /rpt_family="L1"
complement(7767..8037)
rpt_family="ALU"
repeat_region complement(8186..8472)
/rpt_family="ALU"
misc_feature 8473..8625
/gene="WUGSC.H:RC083M05.1"
/note="match to human 3' EST H75782 (NID:g1049794), bases 287-444"
8841..9161
/gene="WUGSC.H:RC083M05.1"
/note="match to human 5' EST H75921 (NID:g1050050), bases 21-348"
9481..9547
/gene="WUGSC.H:RC083M05.1"
/note="match to human 5' EST N22627 (NID:g1130501), bases 276-343"
misc_feature complement(12612..12907)
/rpt_family="ALU"
misc_feature 13670..13793
/gene="WUGSC.H:RC083M05.1"
/note="match to human 5' EST H41382 (NID:g17434), bases 143-266"
143-266"
repeat_region 13794..13877
/rpt_family="ALU"
misc_feature 13878..13906
/gene="WUGSC.H:RC083M05.1"
/note="match to human 5' EST H41382 (NID:g17434), bases 30-58"
repeat_region 13907..14104
/rpt_family="ALU"
complement(14110..14137)
/rpt_family="L1"
repeat_region complement(15618..15907)
/rpt_family="ALU"
repeat_region 17227..17522
/rpt_family="ALU"
18667..19235
/note="match to human fetal brain 5' EST D61494 (NID:g970409), bases 1-255, and to human 3' EST R07476 (NID:g759399)"
19550..19670
/rpt_family="ALU"
21507..37303
/note="Similarly to various SS-RNA virus polyproteins; pseudogene; region of matches and close matches to multiple human ESTs, see R68740 (NID:g842257)"
37316..37489
/note="Grali prediction, score = 80"
evidence=not_experimental
complement(38938..39224)
/rpt_family="ALU"
39225..39707
/note="match to multiple human ESTs, see N30113 (NID:g1148633)"
39800..40085
/rpt_family="ALU"
repeat_region complement(40247..40538)
/rpt_family="ALU"
repeat_region complement(40632..40924)
/rpt_family="ALU"
repeat_region complement(42283..42891)
/rpt_family="ALU"
repeat_region complement(45674..45613)
/rpt_family="ALU"
misc_feature complement(45614..45737)
/note="match to human 3' EST H48898 (NID:g988738), bases 129-333"
misc_feature complement(46107..47026)
/note="match to multiple human ESTs, see N81064 (NID:g1243765), H48897 (NID:g988737), and M78831 (NID:g273146)"
repeat_region complement(47027..47318)

```

Query Match	Best Local Similarity	Matches 1133; Conservative	Score 1035.2; DB 9; Length 56093; Pred. No. 6.1e-304; Mismatches 111; Indels 13; Gaps 2.
QY 1	tcaaatcgaaagagctttagcttctgaacgcgcaaaagaggagggaacctgtttatctttt 60	1	
DB 37018	TCMAAATCGAAGAGCTTTAGACTTGTCTAACCCCTGAAAGAGGGGGAACCTGTTATTITTT 37077		
QY 61	aggggagaagaatgcgtttagtattgtatctaatcaatcctgnaalcatctactgagagaagttaaaga 120	61	
DB 37078	AGGGGAGAAGTCTGTTATTATTGTTAATCAATCCGGAATCGTCACTGAGAAAGTTAAAGA 37137		
QY 121	aattgagatgcaatataatgttagagcagaggaaccttcaaacactgtgacccctggggcct 180	121	
DB 37138	AATTGAGATTCAGAAACAACTGAGAGCAAGAGAGCTTCCAAACACTGAGACCTGGGGCCT 37197		
QY 181	cttcagcaaatgagatgccccttggaacctccctctttagaagacctttagagcctataatatt 240	181	
DB 37198	CCTCAGCCCAATGAGATGCCCTTGATTTCTCCCTCTTTCAGGACCTCTAGCAAGCTATAATATT 37257		
QY 241	ttaactcctctttggaacctgtatcttcaacttccctgtttagtctgttctcttcagaat 300	241	
DB 37258	GCTACTCGCTCTTTGGACCTGTATGCTTTAACTCACTGTTTAACTTTGTCTCTTCAGAAAT 37317		
QY 301	tgaagctgtaaaagctacaataagcttcttcaaatgaaaccccgagatgagtcacatgactaa 360	301	
DB 37318	CGAAGCTGTAACATA-----CAATATGAGGCCCAAGATGACATCCCAAGACTAA 37365		

361 aatcaccgctggaacccctggaacccgctgctagactatgctctgtaattgacattga 420
|||||
37366 GATCTACCGCAGACCCCTGGACCGGCTGTAGCCACGACTGTGAATGATCAATCA 37425
421 agtcacccctcccggaatactcaatgacacccctactaactcaattcaatga 480
|||||
37426 AGGCACCCCTCTGAGAAATCTCAGCTGCACACCTTACTACGCCCAATTCAGCAGG 37485
481 aagcagcttagaagcaatctgacgaacccctcccaacagtaactggtcttcctgttga 540
|||||
37486 AAGCAGTTAGAGCGGTCTGCGGCCAACTCCCAACAGCAGCTAGTTCTCTGTTGAGA 37545
541 gggctgactgagaagacaggaactgactgtatctctaggctgaactaagaaatccaaagcct 600
37546 TGGGGGACTGTAGAGACAGAGACTGATGATTCCTAGGCTGACTAAGATCCCTTAAGCT 37605
601 auctggagagtgaccgcatcatctttaacatgaggcttgcaactagctcaacaccg 660
37606 AGCTGGAAAGTGACACACTTCACTTAAACAGGGGCTTGCACTTAGCTCACACCTG 37665
661 accaatcagaagctcactaataatgctaaatcaggaacaaagaggaagcaatagcc 720
37666 ACCAATTCAGAGAGCTCAGCTAAATGCTAATTAGCAAGAGAGGTAAGAATATAGCC 37725
721 aatcattcatctgctgagagacagcgaggaaagaaagattggagataaaactaagca 780
37726 AATCATCTATTGCTCTGAGAGCAGACAGAGAGGAAATGCGGATTTAAACCCAGTC 37785
781 ttcaagcagcaagacagacacccctctgggtccctccatctgtaaggagctcgtttt 840
37786 TTGAGAGCCGCGCAAGCAGACCCCTTTGGGTCCCTCCCTTGTATGGAGCTGCTTTT 37845
841 caactatctcaactcattaaatcaatcagactgacactctctgtgcgtgttttattg 900
37846 CATGCTATTCTACTCTAATTAATCTTGCAACGCACTCTTGAGTCCATGTTCTTAAGG 37905
901 ctcaagctgagcttctgtctgacacacacactgtctgttgcaacgctcaagaccgct 960
37906 CTGAGCTGAGCTTTCGCTGCGCATCCACACCTGCTGTTGGCGCCAGCAGACCCGCC 37965
961 gctgactcaccatccctctggaatcagcagagtgctcactgtctctgatacgaagcagt 1020
37966 GCTGACTCCCATCCCTCTGAGATCAGCAGGGGTGCTCGCTGCTCTATCAGGAGGC 38025
1021 accaatgccaactcccgatcaggtctaaaggtctgcaattgtctctgcaatgagtc 1080
38026 ACCCATTCGCCGCTCCCAATCGGGCTAAAGGCTTGCCATTTGCTGCAATGGCTAAGTCC 38085
1081 tgggtctgcttaataagactgaacatgacatgacatgacatgacatgacatgacatgac 1140
38086 TGGGTTCATCTCTAATTTAGCTGAACTGACCTGAGGTTCATGAGTCTCTCTGTGAC 38145
1141 ccagagctcttaataagactaataacaccacacgcaatgacccaagaattccatccctgt 1200
38146 CCACAGCTTTAATATAGACTAATACATCCACGATGGCCCAAGTTCCATTTCCCTT-GAA 38204
1201 tctgtgagagccaagaaccccgagtgagaangtgagagcttgccacaaattggagaag 1257
38205 TCCATTAAGGCCAAGAACCCAGGTCAGAGAACAGAGGCTTGCCACCATTTGGGAG 38261

RESULT 5
AC007566/c 149194 bp DNA linear PRI 06-FEB-2002
LOCUS AC007566 Homo sapiens chromosome 7 clone CTB-10G5, complete sequence.
DEFINITION AC007566
ACCESSION AC007566
VERSION AC007566.2 GI:11181861
KEYWORDS HTC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 149194)

AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 149194)
TITLE Waterston, R.H.
JOURNAL Direct Submission
SUBMITTED (15-MAY-1999) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
3 (bases 1 to 149194)
TITLE Waterston, R.
JOURNAL Direct Submission
SUBMITTED (02-OCT-2000) Department of Genetics, Washington
UNIVERSITY, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 149194)
TITLE Waterston, R.H.
JOURNAL Direct Submission
SUBMITTED (16-NOV-2000) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 149194)
TITLE Waterston, R.H.
JOURNAL Direct Submission
SUBMITTED (03-JUN-2002) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 149194)
TITLE Waterston, R.H.
JOURNAL Direct Submission
SUBMITTED (06-FEB-2002) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Nov 16, 2000 this sequence version replaced g1:4835815.

COMMENT ----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web Site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: H.RG010G05

FEATURES
Source 1..149194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="CTB-10G5"
BASE COUNT 45439 a 29540 c 31229 g 42986 t
ORIGIN

Query Match 77.9%; Score 1035.2; DB 9; Length 149194;
Best Local Similarity 90.1%; Pred. No. 6.2e-304;
Matches 1133; Conservative 0; Mismatches 111; Indels 13; Gaps 2;

61 aggggaagaatgctgtatgataatcaatctggaaatctctgagaagattaaaga 120
84926 AGGGGAAGAATGCTGTATATGTTAATCAATCCGGAATCTCACTGAGAAATTAAGA 84867
121 aattgagatgaatataatgacagagaggaactcaaacactgacacctggagcct 180
84866 AATTGAGATGAAATACACTAGAGCAGAGAGAGCTTGGAACACTGAGACCTGGGCGCT 84807
181 cctcagccaatgagatgcccctggaactccctctcttagagaccttagagactataat 240
84806 CCTCAGCCAATGAGATGCGCTGAGATTCTCCCTCTTAGAGACCTTAGAGAGCTAATATAT 84747
241 ttactcctcttggaacctgatactcaactctcctgttgaagttgtctctccagaat 300

|||||
Db 84746 GCTACTCCCTTTGGACCTGTATCTTTAACTCCTTTGTAACCTTGTCTTCCAGAA 84687
Qy 301 tgaagctgtaaaagctatacaatagttcttcaaatgaaaccccgatgcatcatgacta 360
Db 84686 CGAAGCTGTAAACCTA-----CAATGTGAGCCCAAGATGCACTCCCAAGACTAA 84639
Qy 361 aatctacccttgagacccttgaccgagctgctgtagactatgctgtagttaaagactga 420
Db 84638 GATCTAACCGCAACCCCTTGACCGGCTCTAGCCACAGATGTGATGATTAAAGACATCA 84579
Qy 421 agtaccctcccgagaaatctcaactgacaaacccctactaactccaattcagtagg 480
Db 84578 AGGCAACCCCTCCGAGGAATCTCAAGCTGCACAACTCTACTAGCCCAATTCACACAG 84519
Qy 481 aagcagttgagcagttgtcgaagccaaaccccccagaagtaacttggttttctgttgaa 540
Db 84518 AAGCAGTTAGAGCGGTGCTGGCCAACTCCCAACAGCAGCTTAGGCTTTTCTGTGAGA 84459
Qy 541 ggtgagctgagagagagactagctggaattccctagagctgagtaaaatccnaagcct 600
Db 84458 TGGGGGACTGAGAGACAGACTAGCTGATTTCTAGGCTGACTAAGAAATCCCTAAGCCT 84399
Qy 601 anctggaaagtgagccgcatccactttaaagatggggcttgcaacttagctcacaccg 660
Db 84398 AGCTGGAGAGTGACACACATCCACCTTTAAACAGGGGCTTGCAACTTAGCTCACACCTG 84339
Qy 661 accaatcagagagctcactcaaaatgtaatacagcgaagagagtaagaatagcc 720
Db 84338 ACCAATCAAGAGCTCACTAAATGCTAATAGGCMAAAGAGAGGTAAAGAAATAGCC 84279
Qy 721 aatcatctatgctcagagcagaagcagggaaagaaagattgggaataaaactcagga 780
Db 84278 AATCATCTATTGCTGTAGAGGACAGAGAGAGGACAAATATGGGATATAAACCCCAAGTC 84219
Qy 781 ttcaagcagcaaaacagcaaccccttgggtccctccatgtatggagctggtttt 840
Db 84218 TTCCAGCCCGCAACGCAACCCCTTTGGGTCCTCCCTTTGATGAGAGCTCTGTTTTT 84159
Qy 841 caactaatctactaatcaaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 900
Db 84158 CATGCTATTTCATCTATTAAATCTTGCAACTGCACTCTTCTGCTGCTCAATGTTTTCAG 84099
Qy 901 ctcaagctgagcttctgtcgcacatccacacacacacacacacacacacacacac 960
Db 84098 CTTCAGAGCTTTCGCTTTCGCGCATCCACACAGCTGTTTGGCGGCAACCGCAAGACC 84039
Qy 961 gctgagcttcaatcccttggagcagcagagtgctcactgctgctcctgatacagcag 1020
Db 84038 GCTGACTCCCATCCCTCTGATCATGATGAGGTGTCGCTGCTGCTGATTCAGCGAGGC 83979
Qy 1021 accaatgccaatcccgagcagcagcagcagcagcagcagcagcagcagcagcagcag 1080
Db 83978 ACCCAATGCCGCTCCCAATCGGGCTTAAAGGCTTGGCCATGTTCTCTCAATGAGCTA 83919
Qy 1081 tgggttctgctaataagaactgaacactgctcactgctgctgctgctgctgctgctg 1140
Db 83918 TGGGTTCAATCAATGAGCTGAACACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 83859
Qy 1141 ccaagcctctataagactataaactcaacacacacacacacacacacacacacacac 1200
Db 83858 CCACACCTCTTAATATAGCTATTAACACTCACCGCATGGCCCAAGGCTTCATCTCTT -GAA 83800
Qy 1201 tctgtgagcgaagaaaccccgagctcagagaangtggagcttgcacacattgggaag 1257
Db 83799 TCCATTAAGGCCAAGAACCCCAAGGTCCAGAGAACACGAGGCTTGCCACATCTTGGAG 83743

RESULT 6
AX007980 AX007980 10499 bp DNA linear PAT 06-SNP-2000
LOCUS DEFINITION Sequence 3 from Patent WO967395.
ACCESSION AX007980

VERSION AX007980.1 GI:9995677
KEYWORDS
SOURCE
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 10499)
TITLE Perin,J.P., Rieger,F. and Alliel,P.M.
Nucleic sequence and deduced protein sequence family with human
endogenous retroviral motifs, and their uses
JOURNAL INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER
FRANCOIS (FR); ALLIEL PATRICK M (FR)
location/Qualifiers
FEATURES
source 1..10499
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 3048 a 2676 c 2280 g 2495 t
ORIGIN
Query Match 77.0%; Score 1023.2; DB 6; Length 10499;
Best Local Similarity 90.1%; Pred. No. 2.7e-300;
Matches 1132; Conservative 0; Mismatches 111; Indels 14; Gaps 3;
Qy 1 tcaaaatcgaagagcttagacttgctaacgcgcaaaagaggggagactgtttatttt 60
Db 9018 TCAAAATCCAGAGGCTTTGACTGTGTAACCTGAAAGAGGGGAACCTGTTATTATTTT 9077
Qy 61 aggggaagatgctgttagtatgtaatacatctggaatcatcatcactgagaaagtaaga 120
Db 9078 AGGGGAAGATGCTGTATTATGTTAATCAATCCGAATCGACATGAGAAAGTTAAAGA 9137
Qy 121 aattgaagatcgaataatgtagagcagagagacttcaaaacactgcacccctgggct 180
Db 9138 AATTGAGATCGAATTAACAAGTAGAGCAGAGAGCTTCCAAACACACGACCTGGGGCT 9197
Qy 181 cctcaagcaatgtagctccttgagcctcctcctcttagagcctctgagcagcataatct 240
Db 9198 CCGACACCAATGATGCCCTGGATTTCTCCCTTTAGAGACCTTGACAGCTAATATT 9257
Qy 241 ttactcctctcttgagccctgtatcttcaactcctctgttaagttgtctctccagaat 300
Db 9258 GCTACTCCCTTTGGACCCCTGATCTTAACTCTTGTAACTTGTCTTCCAGAAAT 9317
Qy 301 tgaagctgtaaaagctatacaatagttcttcaaatgaaaccccgatgcatcatgacta 360
Db 9318 CGAAGCTGTAAACCTA-----CAATGTGAGCCCAAGATGCACTCCCAAGACTAA 9365
Qy 361 aatctacccttgagcccttgaccgagcctgtagactatgctgtagttaaagactga 420
Db 9366 GATCTAACCGCAACCCCTTGACCGGCTCTAGCCACAGATGTGATGATTAAAGACATCA 9425
Qy 421 agtaccctcccgagaaatctcaactgacaaacccctactaactccaattcagtagg 480
Db 9426 AGGCAACCCCTCTGAGGAATCTCAGCTGCACAACTCTTAACGCCCCAATTCACACAG 9485
Qy 481 aagcagttgagcagttgtcgaagccaaaccccccagaagtaacttggttttctgttgaa 540
Db 9486 AAGCACTTAGAGC -GGTCTCGGCCAACCTCCCAAGAGCACTTAAGGTTTCTGTTGAGA 9544
Qy 541 ggtgagcgaagagagactagctggaatttccctgagctgagtaagaatccnaagcct 600
Db 9545 TGGGGGACTGAGAGAGAGACTAGCTGATTTCTAGGCTGACTAAGAAATCCCTAAGCCT 9604
Qy 601 anctggaaagtgagccgcatccactttaaacaatggggcttgcaacttagctacacccg 660
Db 9605 AGCTGGAGAGGTGACCACTTAAACAGGGGCTTGGCAACTTAGCTACACCTG 9664
Qy 661 accaatcagagagctcactcaaaatgtaatacagcgaagagagtaagaatagcc 720
Db 9665 ACCAATCAAGAGCTCACTAAATGCTAATTAGGCMAAAGCAGGAGTAAGAAATAGCC 9724

QY 721 atcatatattgctgagacacagcggaagaaagattgagatacataacacagca 780
|||||
Db 9725 AATCATATTGCTGTGAGACACAGCAGGAGGCAATGATGCGGATATTAACCCAGTC 9784
QY 781 ttcaagcagaaagacagaccccccttggtgccccccatctgtaigtgagcttgctt 840
|||||
Db 9785 TTTCAGCCGGCAGGACGCCCTTTGGGGCCCTCCCTTTGATGAGAGCTGTGTTT 9844
QY 841 caactatattcaactatlaaatcatgcaactgacactcttcgtgctcgtttttatcg 900
|||||
Db 9845 CATGCTATTTCATCTATTAATCTTGCAACTGCACCTTCTGTCATGTTCTTTACGG 9904
QY 901 ctcaagctgagcttttctgacacacacacacacacacacacacacacacacac 960
|||||
Db 9905 CTTCAGTGTGACCTTCCCTGCGCCATCCACCACTGCTTTGGCCGCCACCGAGCCGCC 9964
QY 961 gctgaactcaccctccttgatgacacagagtgctcactgctcctcgcacacacagag 1020
|||||
Db 9965 GCTGACTGCCATCCCTCTGATGATCGAGGGGTGCTCCCTGCTCTGATCGACGAGGC 10024
1021 acccatgac 1080
|||||
Db 10025 ACCCATGCGCGCTCCCATGCGGCTTAAGGCTTGCATGTTCTGCTGATGCGTAAAGTGC 10084
QY 1081 tgggttgccttaataagaactgaactgctgactggttgcacagcttctcttccacagac 1140
|||||
Db 10085 TGGGTTCATCTTAATTAAGTGAACACTAGTGGGTTCACAGTTCCTCTTGTGAC 10144
QY 1141 ccaagctcttaataagactaatacactcaccgcatgcccagaatccttccctgta 1200
|||||
Db 10145 CCACAGCTTCTTAATGAAGCTATACACTCACCGCATGCGCCACAGTTCATCTTCTT-GAA 10203
QY 1201 tctgtgagcgaagaac 1257
|||||
Db 10204 TCCATTAAGGCCAAGACCCACAGTCAAGAGAACAGAGGCTTGCACACATCTTGGGAG 10260

RESULT 7
AC009727/c 188919 bp DNA linear HTG 16-OCN-2001
LOCUS Homo sapiens chromosome 12 clone RP11-533J15, WORKING DRAFT
DEFINITION
SEQUENCE 5 unordered pieces.
AC009727
VERSION AC009727.5 GI:16118017
KEYWORDS
SOURCE HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 188919)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbarella,J.,
Benton,D., Bimberg,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burck,P., Burrell,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Deen,A.L., Ding,Y., Dinh,H.H., Douhalwaite,K.J., Drepper,O.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gadisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Huliy,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,J., Liu,W.,
Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Martinez,E., Massey,E., Mawhinney,E., McLeod,M.P., Meador,M.,
Mei,G., Metker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenwo,S.,
Ogih,M., Okwuonu,G., Oragune,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Prims,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojuben,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitaishvili,N.,
Stinson,I., Sodergren,E., Sonalike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Washington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 188919)
Worley,K.C.
Direct Submission
Submitted (30-AUG-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 14, 2001 this sequence version replaced gi:12083977.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMIH
Center clone name: RP11-533J15
----- Summary Statistics
Sequencing vector: Plasmid; M7789
Sequencing vector: M13; L08821
Chemistry: Dye-primer; Body: 2% of reads
Chemistry: Dye-terminator; Big Dye; 98% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 186874 bases at least Q40
Consensus quality: 189446 bases at least Q30
Consensus quality: 190585 bases at least Q20
Estimated insert size: 189222; sum-of-coverage estimation
Quality coverage: 8.2x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 61633: contig of 61633 bp in length
* 61634 61733: gap of unknown length
* 61734 133017: contig of 71284 bp in length
* 133018 133117: gap of unknown length
* 133118 164866: contig of 31749 bp in length
* 164867 164966: gap of unknown length
* 164967 180052: contig of 15086 bp in length
* 180053 180152: gap of unknown length
* 180153 188919: contig of 8767 bp in length.
Location/Qualifiers
1. 188919
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-533J15"
56774 a 39075 c 38546 g 54115 t 409 others

FEATURES
source
BASE COUNT
ORIGIN

Query Match 75.7%; Score 1006.6; DB 2: Length 188919;
 Best Local Similarity 86.7%; Pred. No. 3.2e-295;
 Matches 1160; Conservative 0; Mismatches 157; Indels 21; Gaps 4;

QY 1 tcaaaatcgaagagctttagacttgaacttgaacccaaagaaggaggaactgttatttt 60
 Db 22917 TCAAAATCGAAGAGATTATTAAGTGGTAAACACCAAGAGGGGGAACCTGTTATTTT 22858
 QY 61 aagggaagaatgctgttagtagtaagtaacatctggaatcaatcagaagaatlaaga 120
 Db 22857 AGGGGAACATGCTGTATATATGTTAATCAATCTGATCATCAGGAGAGATTAAGT 22798
 QY 121 aatttagatcgaataaagtagaagcaagagagccttcaaaacacbpacccgggggct 180
 Db 22797 AATTCAAGATCGAATACATAGAGAGAGAGGAGGCTTCAAAACACAGGAGCTGGGGCT 22738
 QY 181 cctcaagccaatgagatgccccgagactcctccctcttagagacctcagaataatatt 240
 Db 22737 CTTCAGCCATGAAACACTGTGATTCCTCCCTCTTAGAGACCTAGCACCTAATATT 22678
 QY 241 ttaactcctcttggagacccgttcttcaactcctcgttgaattgtctctccgaat 300
 Db 22677 GTCACTCCTCTTGGACACTGTAACTTAACTCTGTTAAGTTGCTCTCCAGAAAT 22618
 QY 301 tgaagctgtaagaactiaaataagttcttcaaatggaacccagatgagtcacatgactaa 360
 Db 22617 TGAAGCTGTAAAGCTTACAAATGTTCTTCAAAATGAGCCTCAAGATGACATGACTAA 22558
 QY 361 aatctacccgtgagacccctgagacccgctgttagactatgctctgattgaatgacattga 420
 Db 22557 AATCTGGCCGGGAGACCCCTGGAGACGCTGCTAGCCCATGTGTCGATGTTAATGACATCA 22498
 QY 421 agtcaacccctccggaggaaatctcaactgcaaacacccctactacatccaatcagtagg 480
 Db 22497 GGGGACCCCTCGTAGAGAAATCTCAACTGACAAACCCCTACTATGCCCGAGTTCCAGAGG 22438
 QY 481 aagcagtttagagcagttgttagcgaacaccccccacagttacttgggttctcgtttaga 540
 Db 22437 AAGCAGTTAAGTGTGATGTCATGCGCAACCTCCCAACAGCACTTGCGTTTCCTGTTAGA 22378
 QY 541 ggggtgagctgagagagagactgtagattctcctagagctgactaagaatcccaagcct 600
 Db 22377 GGGGAGATCTAGAGACAGAGATTAGTTTCTAGGCGGACTAAGAAATCCCTAAACCT 22318
 QY 601 anctgggaaggtgacgcacatcctttaaactgggggttgcacttaagctacacccg 660
 Db 22317 AGCTGGGAAGGTATCATCATCCACTTAAACATGGGGCTTGCAACTTACTCACACCCG 22258
 QY 661 accaatgagagagctcact-----aaatgctaacagcaaaaacaggaagtagaag 712
 Db 22257 ACCAATAGAGTAGTAAGAAAGACAGCTCACTAAAGCTACTTTGGCAAAAACAGAGAGTAAAG 22198
 QY 713 caataagcaatcatctatctgcttggagagcagcggaagagacaagattggagataaaa 772
 Db 22197 AAATAGCCATTCGCTATGCTGTAGAGACAGAGGAGGAGCAATGATGATATATAA 22138
 QY 773 ctcaagatcagaagccagcaaacagc-aaccccttgggttccctcccatgtatggag 831
 Db 22137 CCCAGGATTCGAGCCGCAATGCTTAAACCGCTTGGGTCCTCCCTTGTATGGGAG 22078
 QY 832 ctctgtttcaactctatctatctatctatctcaactgacactctctgctgcgtgt 891
 Db 22077 CTCGT-----TTCACTATTAATCTTGCACACGCAATCTCTT--CTGGTCC 22030
 QY 892 ttattatgctcaagctgagcttctgttgcacatccacacactgctgtttgccaagctaac 951
 Db 22029 GTGTATAGGCTCTAGCTGAGGCTTTCGCTCCGCAATCCACACTGCTGTGTCGCTGTCG 21970
 QY 952 agaccgctgctgacttccatcccttggatccagcagagaggtcactgctcctgctatc 1011
 Db 21969 AGACCCGCCACTGACTTCATTCCTGTGGGTGGGAGGAGGTCCGGCTGTCCTGTATC 21910

QY 1012 cagcagagtagcccatctgcccacccgatcagcagtaagaagcttgcattgttctgcatg 1071
 Db 21909 CAGCCAGAGCACCATTCCTCTCTCGGATCGGGCTTAAGGCTTGCCATTTGCTCGACGG 21850
 QY 1072 ctatgctcctgggttctgtcctaataagactgaacactgctgactgggttccatgattctc 1131
 Db 21849 CGAAGTCCCTGGGTTCTCTCTAATGACGCTGAATACAGTACATGCTGGTTCACAGGTTCTC 21790
 QY 1132 ttccatgaccccaagcgtcttcaatagagctataaactcaccgcatggtgccaattccat 1191
 Db 21789 TTCATGACCCATGCTTCTTAATAGACTTAAACACTCACCGCATGGCCCAAGATTCAT 21730
 QY 1192 tccctgtatctgtgaagcacaagaacccaggtcagaagaangtgaagcttgcacacatt 1251
 Db 21729 TCCTGGAATCTGTGAGGCGCAAGACCCAGGTCAGAGAACAGAGAGCTTGCCACATCT 21670
 QY 1252 ggggaagtggcccaactgctgatttggtagcgccacccacatcttgggaagctgtagagc 1311
 Db 21669 TGGAGTGGCCACACCATCTTGAAGTGCGCTGCGCACCATCTTGGAGACTCGGGAGC 21610
 QY 1312 aaggaatcccccagtaaca 1329
 Db 21609 AAGGAAGTGGCAGTAGCA 21592

RESULT 8
 AC092708/c 161571 bp DNA linear HTG 20-JUL-2001
 LOCUS Homo sapiens chromosome 18 clone RP11-141A18 map 18, WORKING DRAFT
 DEFINITION
 SEQUENCE: 3 unordered pieces.
 AC092708
 AC092708.1 GI:14971454
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 161571)
 Homo sapiens chromosome 18, clone RP11-141A18
 Unpublished
 2 (bases 1 to 161571)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
 Anderson,S., Barina,N., Bastien,V., Boguslavsky,L., Bouhagalter,B.,
 Brown,A., Camarato,J., Campoliano,A., Chang,J., Chazaro,B.,
 Choquel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
 Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
 Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-pierre,N.,
 Hages,B., Harford,A., Horton,L., Hulme,M., Iliev,I., Johnson,R.,
 Jones,C., Kamet,A., Karatas,A., Kells,C., Labrecque,K.,
 Lamazares,R., Landers,T., Lehoccky,J., Levine,R., Liu,G.,
 Maclean,C., Macdonald,P., Major,J., Margus,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrum,J.,
 Menous,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Risse,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,
 Seaman,S., Severy,P., Sounguez,C., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S.,
 Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zembek,L., Zimmer,A. and zody.M.

TITLE
 Direct Submission
 Submitted (20-JUL-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: L12749
 Center clone name: 141A.18
 Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 160833 bases at least Q40
 Consensus quality: 161248 bases at least Q30
 Consensus quality: 161352 bases at least Q20
 Insert size: 158000; agarose-1p
 Insert size: 161371; sum-of-ctrls
 Quality coverage: 9.8 in Q20 bases; sum-of-ctrls
 Quality coverage: 9.6 in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 32930: contig of 32930 bp in length
 * 32931 33030: gap of 100 bp
 * 33031 114301: contig of 81271 bp in length
 * 114302 114401: gap of 100 bp
 * 114402 161571: contig of 47170 bp in length.

FEATURES

source
 1. 161571
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="18"
 /map="18"
 /clone="RP11-141A18"
 /clone_1b="RPC1-11 Human Male BAC"
 1. 32930
 /note="assembly-fragment
 clone.end:SP6
 vector.side:left"
 33031..114301
 /note="assembly-fragment"
 114302..161571
 /note="assembly-fragment
 clone.end:T7
 vector.side:right"
 E COUNT 45518 a 33859 c 34317 g 47677 t 200 others
 ALIGN

Query Match 74.5% Score 989.6; DB 2; Length 161571;
 Best Local Similarity 88.6%; Pred. No. 4.9e-280;
 Matches 1146; Conservative 0; Mismatches 117; Indels 30; Gaps 6;

QY 1 tcaaaatcgaagagcttgaacttgaaccgcaaaagaggggaactgttattttt 60
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 23281 TCATAAATTGAAGAGCTTGAAGCTTGAACCAACCAAGAGGGAACCTGTTATT 23222

QY 61 aggggaagaatgctgttagtctgttaatacctcgtgaatcatctcctgagaagttaa 120
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 23221 AGGGGAAGAATGCTGTATTATTGTTAAATCAATCCGGAATCATCCGGAAGATT---A 23166

QY 121 aattgagatgaataataatctagaagagagagagagagagagagagagagagag 180
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 23165 AATTCAAGATCAAAATACATATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 23106

QY 181 cctcagcaaaagagatgacctcctcctcctcctcctcctcctcctcctcctcctcct 240
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 23105 CCTCAGCCAAATGAGATGCGCTGATTCCTCCTCTTGTAGACCTTACAGAGCTATATATT 23046

QY 241 ttatctcctcttgagaccctgtatctcaactcctcctgttaagttgtctctccagaat 300

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 23045 GCTACTCTCTTTGGAGCCCTGTATCTTAACTCTCTGTTAAGTTGTCTCTCCAGAT 22386

QY 301 tgaagctgaagagcttaaaatagttcttcaatggaaccagaatgagctcagactaa 360
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 22985 TGAAGGTGAAGCTACAAATCAATCTTCANATGAGGCCAGATGCACTCATGACTAA 22926

QY 361 aatcaccgttgagaccctcgtacgagcctcgtacagactaactcgtatgttaatgaatga 420
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 22925 AATTCACCTGTGAGCCCTGAGACAGAGCTGCTAGGCCATGCTCCAAATGTATGACATCA 22866

QY 421 agtaccctcctcgaaggaatctcaactcgaacacccctactacactcctcaatgaatga 480
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 22865 AGGACCCCTCCCAAGGAATCTCAACGCAACACCCCTTACACCCCAATTCAGCAGG 22806

QY 481 aagcagttgagagagctgtcagcgaacacctcccaacagtaactgtgttctcgttga 540
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 22805 AAGCAGTTAGAGCAGTTGTACCCCAACACTTGGGTTTCTCTGTGAGA 22746

QY 541 ggtgagctgag 600
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 22745 GGGGGAGACTGAGAGACAGAGACTAGCTGTGATTTCTAGAGCCAAATGAATCCCTAGGCT 22686

QY 601 aactggaaggtgacgcgcatctccttaaacatgagggtctgcaactagctcaacccg 660
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 22685 AGCTGGGAAGGTGACCGCATCCACTTTAAACATGGGGCTTGCAATGCTACACACCGG 22626

QY 661 ac-----caacagagagctcactaaatgtcctaatcagagcaaaacagagagaa 712
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 22655 CCAACAGGTAAATTAAGAGAGACTCACTAAATGCTAATAGGCAAAAACAGAGTAAA 22566

QY 713 caatagcaatcatctatctgtcgtgagagacagcggagagagagagagagagagagat 772
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 22565 AAATAGCCAAATCAATCTTTCGCTGAGACACAGTGGGGGGAATGATCAGATATATA 22506

QY 773 ctcaagcatcgaagcagagcaaacacccctcttggtctccctcccatctgatggagc 832
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 22505 CCCAGGCAATTAACCCCAACCAATGCTACCTTTGGTCCCTCTCTTGTATGGGAGC 22446

QY 833 tctgttcaactctatctcaactcactaaacatgagcaact--ggaactctctgtctcgt 890
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 22445 TCTGT-----TTTCGCTATTAATTAATCTTGCAACTGCGCTCTCTGTGGTGCATG 22396

QY 891 tttttat-----gctcaagctgagcttctgtctgcacatccacactgctgttgcac 945
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 22395 TTTGCTATGCTCGGCTCTAGCTGAGCTTTGTTACATGTCACACAGCTGTTTGGCTGC 22336

QY 946 cgtcaagagaccgctgctgacttccatccctcttgatccagagagtgctcaactgtgtc 1005
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 22335 CGCCGACAGCTGCGCACTGACTTCATCCCTTCGATTCGCGAGGAGTCCCTGTGTGCC 22276

QY 1006 ctgattcagcagagtlaccatctgcaactccagatccagagagagagagagagagagag 1065
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 22275 CTGATCCAGCAAGGACACCATGCTGCTCGTCGATCAGGCTAAAGCTTGCAATGTTCT 22216

QY 1066 gcatgctgaagtgcctggtgttctcctaaatagaactgaacatgctcgtcagtggttcc 1125
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 22215 GCATGGCTTAAGTGGCCAGAGTCAATCTTAATGAGCTGAACCTAGTCACTGGGTT--CA 22157

QY 1126 gtctcttccatgagccagcgtctcctaaatagagctataaactcaacagcagagccaa 1185
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 22156 GTTCTCTTCCTGACCCCTGCTCTTAATAGAGCTATNAACCTCACTCAAGGCCCAAGA 22097

QY 1186 ttcaatcctcgtatctgttgaagcgaagaccccaagctcagagaagctgaagcttgca 1245
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 22096 TTCCATTTCTTGGAATCATATAGGCAAGAACCCAGTCAAGAGAACAGAGGCTTGCA 22037

QY 1246 ccaattgggaagtgcccaatcttgtagc 1280
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 22036 CCAATTTGGAAGTGCGCTGCTGCAATCTTGGAGC 22002

RESULT 9

LOCUS	AC013759	166700 bp	DNA	linear	PRI 29-JUN-2001
DEFINITION	Homo sapiens chromosome 18, clone RP11-2E13, complete sequence.				
ACCESSION	AC013759				
VERSION	AC013759.6	GI:14029981			
KEYWORDS	HTG.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
AUTHORS	1 (bases 1 to 166700)				
TITLE	Birren,B., Linton,L., Nusbaum,C. and Lander,E.				
JOURNAL	Homo sapiens chromosome 18, clone RP11-2E13				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 166700)				
	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heatford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K., lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., MacDonald,P., Major,J., Margulis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mhova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Plunkhank,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Riebeck,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S., Severy,P., Sougnuez,C., Spencer,B., Stange-Thomas,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
REFERENCE	3 (bases 1 to 166700)				
AUTHORS	Anderson,S., Barna,N., Nusbaum,C., Lander,E., All,A., Allen,N., Birren,B., Linton,L., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,M., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goid,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heatford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K., lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., MacDonald,P., Major,J., Margulis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mhova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Plunkhank,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Riebeck,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S., Severy,P., Sougnuez,C., Spencer,B., Stange-Thomas,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (15-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
REFERENCE	3 (bases 1 to 166700)				
AUTHORS	Anderson,S., Barna,N., Nusbaum,C., Lander,E., All,A., Allen,N., Birren,B., Linton,L., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,M., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goid,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heatford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Larocque,K., lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C., MacDonald,P., Major,J., Margulis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mhova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Plunkhank,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Riebeck,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S., Severy,P., Sougnuez,C., Spencer,B., Stange-Thomas,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (29-JUN-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
COMMENT	On May 13, 2001 this sequence version replaced gi:14017510. All repeats were identified using RepeatMasker: Smit,A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html				
TITLE	Genome Center				
JOURNAL	Center: Whitehead Institute/ MIT Center for Genome Research				
COMMENT	Center code: WIBR				
	Web site: http://www-seq.wi.mit.edu				
	Contact: sequence_submissions@genome.wi.mit.edu				
	Project Information				
	Center project name: L2525				
	Center clone name: 2_E_13				
FEATURES	Location/Qualifiers				
SOURCE	1..166700				

```
repeat_region /rpt_family="MER103"
complement(19991..20119)
/rpt_family="MIR"
complement(20181..20673)
/rpt_family="MLT1G3"
complement(21649..21939)
/rpt_family="Alusx"
complement(22028..22245)
/rpt_family="L1MB8"
22383..22658
/rpt_family="MLT1A1"
22689..22776
/rpt_family="MLT1A1"
complement(23084..23320)
/rpt_family="L1ME4A"
23541..23586
/rpt_family="(ATG)n"
complement(24127..24226)
/rpt_family="MIR3"
24560..24769
/rpt_family="MIR3"
25483..26173
/rpt_family="MER21B"
complement(26203..26366)
/rpt_family="MIR"
complement(26379..26779)
/rpt_family="MSTB"
27372..27438
/rpt_family="(CAG)n"
28669..29257
/rpt_family="MER4B"
29324..29465
/rpt_family="MIR3"
29489..29841
/rpt_family="THE1B"
complement(29844..30059)
/rpt_family="MER20B"
complement(30458..30834)
/rpt_family="MER65A"
complement(30935..31258)
/rpt_family="AluDb"
complement(32849..33446)
/rpt_family="T1ger3(Golem)"
complement(33448..33750)
/rpt_family="Alusg"
complement(33778..34403)
/rpt_family="PABL_B"
complement(34432..34481)
/rpt_family="PABL_B"
complement(34482..34531)
/rpt_family="T1ger3(Golem)"
complement(34540..34639)
/rpt_family="T1ger3(Golem)"
34740..34928
/rpt_family="MER58A"

Query Match 74.5%: Score 989.6; DB 9; Length 166700;
Best Local Similarity 88.6%: Pred. No. 4.9e-290;
Matches 1148; Conservative 0; Mismatches 117; Indels 30; Gaps 6;

Db 1 tcaaaatcgaagagcttagactctgacgcaaaagagggagagcttattttt 60
42844 TCANAATTGAAGAGCTTAGACTTGTACCAACAAAGAGAGGAACTTATTATTTT 42903

QY 61 aggggaagatcgtctgtatgtaacatctgaaatcaactaactgaaagttaaga 120
42904 AGGGGAAGAATGCTGTATTGTATTATCATCCGGAATCATCAACCGAAGATT---A 42959

QY 121 aacttgagatcgaatataatctagagcagagagaccttcaaaacactgcacccctggagcct 180
42960 AATTCAAGATCAAAATACAAATAGAGCAGAGAGCTTCAAAACACTGAGACCCCTGGAGCCT 43019

QY 181 cctcagcaatcgaatgccttgactctccctctcttgagaccttagagactataatatt 240
```

```
Db 43020 CCTCAGCCAAATGATGCTCCTGATTTCCCTCTTAGAGACCTAGAGCTAATAATTT 43079
QY 241 tttaactctcttgagaccctgatcttaacttaactctcttgtaagtgtgtctctccaagt 300
43080 GCTACTCTCTTGTGAGACCTCTATCTTTAACTCTCTTGTAAAGTTGTCTCTCCAGANT 43139

QY 301 tgaagctgtaaaagctacaatagttctctcaaatgaaacccaagtgcagtcacatgactaa 360
43140 TGAAGGTATAAGCTACAAATCATTTCTTAATGAGAGCCCAAGATGAGCTCATGACTAA 43199

QY 361 aactacccgtgagacccttgagaccgctcgtctagactatgctctgatttaataatgata 420
43200 AATCTACGCTGAGACCCCTGAGACCAAGCCGTGACCAATGTAATGACATCAAA 43259

QY 421 agtaccctccctccgagaaactcaactgacaaaccttaataactcaactcaatgagta 480
43260 AGGCAACCCCTCCCAAGGAATCTCACTGACCAACCCCTTACACCCCAATTTCAGCAG 43319

QY 481 aagcagttagaagctgtgacgcaacctcccaacagtaacttggtttctctgttga 540
43320 AAGCAGTTAGAGAGTTGTGTAGCAACCTCCCAACAGCAGCTTGTCTGTGAGA 43379

QY 541 ggttgagactgagagacagagactgtaattctctagctgtaactaagaatcccaagcct 600
43380 GGGGGAGCTGAGACAGAGACTACTGTGATTTCTTAGGCCCAATGAAGATCCTAAGCCT 43439

QY 601 aactgggaagtgagaccgacatcatttaaaatgagggctggaacttagtcaaccg 660
43440 AGCTGGGAAGGTGACCGCATCACCCTTTAAACATGAGGCTTGCAACTTAAGCTCACACCG 43499

QY 661 ac-----caatcagagagctcaataaaatgctaactgaaacaaacagaggtaaag 712
43500 CCAACAGGTAAATGAAGAGAGCTCACTAAATGCTAAATCAGGCAAAACACAGAGTAA 43559

QY 713 caatagccaatcatatctatgctctgagacagaggggaagaagaagtgtgatatataa 772
43560 AAATAGCCAAATCATCTTTCGCTGAGACACAGTGGGGGACATGATGCAATATATAA 43619

QY 773 ctacagcatcaaaagcagcaacaaacacccctcttggttccctcccaattgataggagc 832
43620 CCCAGGCAATTAAACGCAACGATGAGCTTACCCCTTTGGGTCCCTCTTTATGAGGAGC 43679

QY 833 tctgttttactctattactactataaatacatgaaact--gactctcttggtccgtg 890
43680 TCTGT-----TTTCGCTTATTAAATCTTGCAACTCGCTCTCTTCTGTGTGATG 43729

QY 891 ttctttat-----ggtcgaagctgagcttgtctgcaatccacacatgctgttgcac 945
43730 TTTCGCTATGCTCGGCTCTAGCTGAGCTTTGTTTCACTGTCACACAGCTGTTTGTCTGC 43789

QY 946 cgtacagaccgctgtctgacttcaatccctcttgatcagcaagagtgtaactgtgtc 1005
43790 CGCGGAGACCTGCGACACTGACTTCATCCATCCCTTGGATTCGCGAGAGGTCTCTGTGCC 43849

QY 1006 ctgatacagagaggtatccaatctccactcccgatcaggtctaagagctgtccatgttcc 1065
43850 CTGATTCAGCAAGGACCCCATTTGCTGCTGATTCAGAGCTTAATCTTTGGCCATGTTCT 43909

QY 1066 gcatgagctaaagctgaggtgtgctcctaataagaaactgaaactgtgtcactgttccatg 1125
43910 GCATGGCTAATGGCCAGAGTTTCATCTTAATTGAGCTGATGACTGTCTGTGATT--CA 43968

QY 1126 gtctctctcaatgacccagctctctaataagatataaactcaaccgcatggtccaa 1185
43969 GTTCTCTTCCCTGACCACTGCTCTTAATAGAGCTAATACCTCACTGCAATGGCCAAAG 44028

QY 1186 ttcaatctctgtatctgtgtaggccaagaaccccaaggtcagagaangtgaggtctgcca 1245
44029 TTCCATTTCTCTTGGAATCATGATGAGGCAAGAACCCCAAGGTCTAGAGAACGAGGCTTGCCA 44088

QY 1246 ccatttgggaagtggtccactgtccatttggtaagc 1280
```


Db 44089 CCATTTTGGAGTGGCCTGCTGCCATCTTGGAGC 44123

RESULT	10
AP001488/c	
LOCUS	AP001488
DEFINITION	Homo sapiens chromosome 18 clone RP11-693N15 map 18q12, WORKING DRAFT SEQUENCE, 12 unordered pieces.
ACCESSION	AF001488
VERSION	AP001488.3
KEYWORDS	GI:1988499
SOURCE	HTG; HTGS-PHASE1; HTGS-DRAFT.
ORGANISM	Homo sapiens DNA, clone:RP11-693N15. Homo sapiens

REFERENCE
AUTHORS
1 (Pages 1 to 186911)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 186,911 genomic DNA of 18q12
Published Only In Database (2000) In press
2 (Pages 1 to 186911)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (17-MAR-2000) Masahira Hattori, The Institute of Physical

COMMENT On Jul 14, 2000 this sequence version replaced gi:8117346

Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: <http://hgp.gsc.riken.go.jp/>
Contact: hattori@gscc.riken.go.jp

```
----- Project Information
Center project name: HumDrafl18
Center clone name: RP11-693N15
----- Summary Statistics
```

Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator; ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 182485 bases at least 40
Consensus quality: 184520 bases at least 30
Consensus quality: 185500 bases at least 20
Insert size: 185611; sum-of-contigs
Quality coverage: 9.41k in 320 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 12 configs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the configs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

38569	6616	38468	3446
1	contig of	contig of	in length
91874	91873	27548	in length
112197	contig of	25657	in length
132536	contig of	20123	bp in length
147361	contig of	20340	bp in length
147361	contig of	14725	bp in length
171550	contig of	17155	bp in length
171550	contig of	6814	bp in length
178451	contig of	6821	bp in length
183109	contig of	4558	bp in length
183210	contig of	3146	bp in length
183250	contig of	456	bp in length
186456	contig of		

* NOTE: This is a 'working draft' sequence. It currently consists of 12 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence

* as soon as it is available and the accession number will
* be preserved.

*	1	38468:	contig of 38468 bp	in length
*	38469	38568:	gap of 100 bp	
*	38569	66116:	contig of 27548 bp	in length
*	66117	66216:	gap of 100 bp	
*	66217	91873:	contig of 25657 bp	in length
*	91874	91973:	gap of 100 bp	
*	91974	112096:	contig of 20123 bp	in length
*	112097	112196:	gap of 100 bp	
*	112197	132536:	contig of 20340 bp	in length
*	132537	132636:	gap of 100 bp	
*	132637	147361:	contig of 14725 bp	in length
*	147362	147461:	gap of 100 bp	
*	147462	164616:	contig of 17155 bp	in length
*	164617	164716:	gap of 100 bp	
*	164717	171530:	contig of 6814 bp	in length
*	171531	171630:	gap of 100 bp	
*	171631	178451:	contig of 6821 bp	in length
*	178452	178551:	gap of 100 bp	
*	178552	183109:	contig of 4558 bp	in length
*	183110	183309:	gap of 100 bp	
*	183310	186355:	contig of 3146 bp	in length
*	186356	186455:	gap of 100 bp	
*	186456	186911:	contig of 456 bp	in length.

FEATURES	Location/Qualifiers
source	1. .186911

```
misc_feature      /organism="Homo sapiens"  
                  /db_xref="taxon:9606"  
                  /chromosome="18"  
                  /map="18q12"  
                  /clone="RP11-693N15"  
1..38468  
/note="assembly_fragment"  
38569..66116  
/note="assembly_fragment"  
66217..91873  
/note="assembly_fragment"  
91974..112096  
/note="assembly_fragment"  
112197..132536  
/note="assembly_fragment"  
132637..147361  
/note="assembly_fragment_clone_end:T7 vector_side:left"  
147462..164616  
/note="assembly_fragment"  
164717..171530  
/note="assembly_fragment"  
171631..178451  
/note="assembly_fragment"  
178552..183109  
/note="assembly_fragment"  
183210..186355  
/note="assembly_fragment"  
186456..186911
```

```
BASE COUNT      55081 a 38409 c 38950 g 53371 t 1100 others
ORIGIN          /note="assembly_fragment clone_end:SP6 vector_side:left
```

Query Match	74.58;	Score 989.6;	DB 2;	Length 186911;
Best Local Similarity	88.68;	Pred. No. 4.9e-290;		
Matches 1148; Conservative	0;	Mismatches 117;	Indels 30;	Gaps 6

Oy	1	tcaaaatcgaagagcttagactgtctaaccgcgcaaaagaggggaaacctgttattttc	60
Db	25843	TCAAAATTCGAGAGCGCTTTAGACTTGCCTAACCCACCAAAGACAAGGAACCTGTTTATT	25784

61 aggggaagaatgctgttagtattaatcaatctgyaatcatctactgagaaagttaaaga 120

Db 25783 AGGGAAGAAATGCTGTATTATGTTAAATCAATCGGAATCATCACCGAAGATT----A 25728

QY 121 aattgagatcgatatataatgttagagcagagaccttcaaacatgcacccctgggcct 180

```

Db 25727 AATTCAGATCAAAATACATGAGAGAGAGAGCTTCAAAACACTGAGACCTGGGGCT 25668
Qy 181 cctcagcaatgagatgacctgagacctccctctcttaagaccttaagacatatatc 240
Db 25667 CCTCAGCCAAATGATGAGCTGATTTCTCCCTCTCTTGAAGACCTCAGACGATATATTT 25668
Qy 241 ttactctctcttgagacctgatactcaactctctgttaagttgtctctccagat 300
Db 25607 GCTACTCTCTTTGGACCTGATATCTTAACCTCTTGAAGTTGTCTCTCCAGANT 25548
Qy 301 tgaagctttaaagctacaaatagttctcaatagaaacccagatgagctgacatgactaa 360
Db 25547 TGAAGGTATAAGCTACAAATCATCTTCAATGAGGCCAGATGAGCTCATGACTAA 25488
Qy 361 aatcactgctgagccccggagccgctcctcagactatgacctgtatgtaagcatgta 420
Db 25487 AATCTACCGTGGACCCCTGGACACCGCTGTAAGCCCATGCTCCAAATGTAATGACATCAA 25428
Qy 421 agtccccctccggaggaatctcaactgcaaacccctactacatcccaattcagtag 480
Db 25427 AGGCAACCCCTCCCAAGAAATCTCAACTGCAAAACCCCTACTACACCCCAATTGACAGG 25368
Qy 481 aagcagtttagagcagttgtctcagccaaactcccaacagtaactgtgttctctgtaga 540
Db 25367 AAGCAGTTAGAGCACTTTCAGCCAACTCCCAACACACTGTGGTTTCTGCTGTGAGA 25308
Qy 541 ggttgagctgagagacagagtaactgtatctcctcagctgacataagatcccaagct 600
Db 25307 GGGGGGACTGAGAGACAGAGACTAGCTGATTTCCTAGGCCAACTAAGATCCCTTAGCCT 25248
Qy 601 anctggaaggttgacgcgacatccatttaacaatgggctgcaactagctacacccg 660
Db 25247 AGCTGGGAAGGTGAGCCGATCCACTTTAAACATGGGGCTTGCAACTTAGCTCACACCGG 25188
Qy 661 ac-----caatcagagagctcaataatgctcaatcagcaaaacagaggttaag 712
Db 25187 CCAACAGGTAATTAAGAGAGCTCACTAAATGCTAATCAGGCANAAACAGAGGTTAA 25128
Qy 713 caatgacacatcatctatgctcagagacagcgaggaagaaagaaatgagatatataa 772
Db 25127 AAATAGCCAAATCATCTTTCGCTGAGACACAGTGGGGGCAATGATAGATATAA 25068
Qy 773 ctcaaggaatcaagcagcaaacaccccttggtgtccctccatctgtagagagc 832
Db 25067 CCGAGGCAATTCAGGACCAATGCTACCTTTGGGTCCCTCTCTTGTATGGAGC 25008
Qy 832 tctgtttcacctcattcacatctataatcatgcaact--gcaactcttggtcgtg 890
Db 25007 TCTGT-----TTTCGCTCTATTAAATCTTGCAACTGGCTCTCTTGTGTGATG 24958
Qy 891 ttttttlat----ggtcaagctgagcttctgtctgcacatccacacatctgttgcac 945
Db 24957 TTTGCTATAGGCTCGGCTCTACTAGCTGAGCTTTGTTCACATGTCACACAGCTGTTTGTGTC 24898
Qy 946 cgtcaagaaacccgctgctgactcattcccttggaatccagcagagvtgctcagctgtc 1005
Db 24897 CGCCGCAACCTCGCAGCTGATCCATCTTCGATTCGCGATTCGCGAGGTCTCTCTGTCGCC 24838
Qy 1006 ctgataccagaggttaacatctgcaactccgatacagcttaagagctgacatgttct 1065
Db 24837 CTGATCCAGCAAGGACCATTTGCTGCTCTGATACAGGCTAAAGCTTGCCATGTTTCT 24778
Qy 1066 gcaatgctaaatgctcgtgtctgtctcctaataagaaactgtaactgtgtctcagctg 1125
Db 24777 GCATGCTAAGGCGCCAGCTTCATCTTAATGAGCTGAACACTAGTCACTGGGT--CACA 24719
Qy 1126 gttctctcctcagacccagctctcctaataagactataacacacagcagctgagccaa 1185
Db 24718 GTTCTCTTCCTGAGCCCTGATCTTAATGAGCTATTAACACTCACTCAAGGCCCAAGA 24659
Qy 1186 ttcacatctctgtatctgtaggagcagaacccaggttcagagaangtgaggtctgca 1245

```

```

Db 24658 TTCCATTCTTGGAAATCCATGAGGCCAAGAACCCAGGCTGAGAGAACAGAGCTTGCCA 24599
Qy 1246 coatttgggaagtggccacactgacatttgtagc 1280
Db 24598 CCATTGTGGAAGTGCGCTGCTGCATCTTGGAGC 24564

RESULT 11
AP001599
LOCUS 118241 bp DNA linear PRI 29-APR-2000
DEFINITION Homo sapiens genomic DNA, chromosome 21, clone: KB126A3, App-D21S292
ACCESSION AP001599
VERSION AP001599.1 GI:7670553
KEYWORDS HTG.
SOURCE Homo sapiens pre-pro-B cell cell_line: FLEB14-14 DNA, clone_lib: Keio
BAC library clone: KB126A3.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 118241)
AUTHORS Shimizu,N., Kudoh,J. and Shibuya,K.
TITLE Homo sapiens genomic DNA, chromosome 21, clone: KB126A3, App-D21S292
region
JOURNAL Published Only in Database (2000) In press
REFERENCE 2 (bases 1 to 118241)
AUTHORS Shimizu,N., Kudoh,J. and Shibuya,K.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2000) to the DDBJ/EMBL/GenBank databases.
Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular
Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan
(E-mail: nshimizu@med.keio.ac.jp, Tel: 81-3-3351-2370,
Fax: 81-3-3351-2370)
FEATURES
source
1..118241
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="FLEB14-14"
/cell_type="pre-pro-B cell"
/chromosome="21"
/clone="KB126A3"
/clone_lib="Keio BAC library"
BASE COUNT 35861 a 22962 c 22601 g 36817 t
ORIGIN

Query Match 74.2%; Score 985.6; DB 9; Length 118241;
Best Local Similarity 88.5%; Pred. No. 8e-289;
Matches 1120; Conservative 0; Mismatches 122; Indels 23; Gaps 4;

Qy 1 tcaaaatcgaagagcttagacttgctaacgccaagaaggggaacgtgtatttt 60
Db 63198 TCAAAATGAGAGAGCTTATGAAATGCTAATCAGAGAGGGGAGAGCTTTTATTTT 63257
Qy 61 aggggaagaatgctgtatgttaatcaatctggaatcattactggaagaagttaaaga 120
Db 63258 AGGGGAAGAATGCTTATTTATGTTAAATCGAATTCGAAATCAACCAAGAAATTAAGA 63317
Qy 121 aattgagatcgaatatatgtagagcagaggaaccttcaaaa--cactgcacccggggcc 179
Db 63318 AATTCAGATGGAATACAGCTAGACAGAGAGAGCTTAAAAAACACTGAGACCTGGGCC 63377
Qy 180 tctcagccaatlgatgctcgtgactctccctctcttagagaccttaagcagataatat 239
Db 63378 TCTCAGGCCAATGATGATGCTGATTTCCCTCTTATGAGACCTCTACAGCT--ATAT 63434
Qy 240 tttaactctctcttggaacccgtatctcaactctcctgttaagtgtgtctctccgaa 299
Db 63435 TTCTACTCTCTTTGGACCCGTATCTTAACCTCCGCTGAAGTTGTCTCTCCACAA 63494
Qy 300 ttgaagctgataactacaatagttctcaaatggaacccagatgagctcagacta 359
Db 63495 TCGAAGATGTAACATCAATCTTCTTCAAAATGAGACCCCAAGATGAGTCAATGACTA 63554

```

QY	360	aactctacccgtagaccccttgagccgcgcctgtctagactctgtctgattgtaataagcatg	419
Db	63555	AGATCTACTGAGGACCCCTGGACACAGCCTGCTAGCCCATGCTCCAAATGTTAATGACATTG	63514
QY	420	aagtcacccctcccgagaaatctcaactgacacacccctctaactcaatccaatccagtag	479
Db	63615	AAGGCAOCCCTCCCAAGGAATCTCACTGACAAACCCCTACTGTCTCAATTCAGAG	63674
QY	480	gaagcaagtagagcagatgtgtcaagccaactccccaacagtaactgtggtttccgtttag	539
Db	63675	GAACCAAGTTACAGTGTCTCTGGCCAACTCCCAACAGCAFTTGTGATTTCTGTTGGG	63734
QY	540	agggtgagctagagagacagcagctagctgtgatttctctgagtcgactaagaatccnaagcc	599
Db	63735	AGGGGGCACTGAGAGACAGGACTAGCTGATTTCTTAGCGTCACTGAAATATCCCTAAACCC	63794
QY	600	tancctggaaggttagccac	659
Db	63795	TAGCTGGGAAGGTAGCACTCTCCACCTTTAAACACAGGCGCTTGCACTTACTTCACACC	63854
QY	660	gaccacac-----agagagctcactaaatgcttaacacaggaacaggaagtaa	710
Db	63855	TACCAATTTGGATAGTAAAGAGAGGTCACTAAATGCTAATTAGGCAAAACAGAGGTAA	63914
QY	711	agcaatagacacatctatctgtcctggaagacacagcggaaggaaggaagattggatta	770
Db	63915	AGAAATAGCCAAATCATCTTCCTATTGCTTGAAGGACAGCGGGAAGGACAAATGACGAGATTA	63974
QY	771	aactcagagcattcaagccacagacagcaacacccctttgtgcctccctcccatgtgatagga	830
Db	63975	AACCCAGGCATTTCCAGCCTGACAGGCAACCCCTTTGGTCCCTCTCTTTGATGGGA	64034
QY	831	gctctgttttcaactctatcttcaactctatlaaatacgaatgcaactgcaactctctgtgcgtg	890
Db	64035	GCTCTGT-----TTTCACCTCTATTCATCTTGCACACTGCACCTCTCTGGTCCGTG	64084
QY	891	ttttttttaggtcgaagcagcagcatttgtgtgcacacacacacacacacacacacacacgca	950
Db	64085	TTTGTATAGGCTCAAGCTGACACTTTTGTCTCACACATCCACACTGCTGTTTCCCGCGTGG	64144
QY	951	cagaaccgcctctgactctcaatccctttggaatccagacagagtgctcaactgtgctcctgat	1010
Db	64145	CAGAACCGTGCCTGACTTCCATCCCTCCAGATCTGAGGAGGAGGTGTCACATGTGCTCCTAT	64204
QY	1011	ccagcagagtagcccaattgccaatcccgatccagctcagcgtaaagcgttgccattgttccatgca	1070
Db	64205	CCAGGAGGACCAACCATTCCTCCACTCCCGATTCAGGCTAAAGGCTTGCCATTGTCTCCGACA	64264
QY	1071	gctaagtgccgtggtgttgttcttaataagaaactggaacacagcagcagcagcagcagcagcagc	1130
Db	64265	GCTAAGTCCCTGGGTTCTCTTAATCAAGCTGGAACACTAGTACCTGAGGCTGCTCATAGTGTCT	64324
QY	1131	cttccatgacccacagcgtcttcaatagagctlaaacacacacacacacacacacacacacacac	1190
Db	64325	CTTCCATGACCCCATGAGGCTTTAATAGACCTTAACACTCACCGCATGGGCCCAAGATTCCA	64384
QY	1191	tttcccttgatctgtgagggccaagaaccccaagcagtgacagaangtgaagctgtgccaacat	1250
Db	64385	TTTCTTGGAAATCCGTGAGGCGCAAGAACCCCAAGGTCTAGAGAACACAGAGCTCCGCGCATCT	64444
QY	1251	tgga 1255	
Db	64445	TGGAA 64449	

RESULT	12
AP001697	
LOCUS	34000 bp DNA linear PRI 30-MAY-2000
DEFINITION	Homo sapiens genomic DNA, chromosome 21q, section 41/105.
ACCESSION	AP001697 AL163242 BA000005
VERSION	AP001697.1 GI:7768705
KEYWORDS	.

SOURCE	Homo sapiens DNA.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (sites)
AUTHORS	Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T., Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Seoda,E., Ohki,M., Takaki,T., Sakaki,Y., Taudien,S., Blechschmidt,K., Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R., Patterson,D., Reichwald,K., Rump,A., Schillabel,M., Schudy,A., Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasak,K., Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S., Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G., Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A., Reichel,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S., Hennig,S., Riesselmann,L., Daggnd,E., Wehrmayer,S., Borzzym,K., Gaspard,K., Mizetic,D., Francis,F., Lehnach,H., Reinhardt,R. and Yasp,M.L.
TITLE	Direct Submition
JOURNAL	Submitted (10-APR-2000) The Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic Sciences Center, Human Genome Research Group * Institute of Molecular Biotechnology, Genome Analysis * Keio University School of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular Genetics (addresses see below)
COMMENT	On May 30, 2000 this sequence version replaced gi:7717295. The chromosome 21 mapping and sequencing consortium consisting of * RIKEN Genomic Sciences Center, Human Genome Research Group, * Sagamihara 228-8555, Japan, * e.mail: hattori@gscl.riken.go.jp * URL: http://hgp.gsc.riken.go.jp/ and * Institute of Molecular Biotechnology, Genome Analysis, * Beutenbergstrasse 11, D-07745 Jena, Germany, * e.mail: gscj-submit@genome.imb-jena.de * URL: http://genome.imb-jena.de/ and * Keio University School of Medicine, Molecular Biology, * Tokyo 160-8582, Japan, * e.mail: nshimizu@db-med.keio.ac.jp * URL: http://www.db.med.keio.ac.jp/ and * GBF, Dept. of Genome Analysis, * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail: info.genome@gbf.de * URL: http://genome.gbf.de/ and * Max-Planck Institute for Molecular Genetics, * Innesstrasse 73, D-14195 Berlin, Germany, * e.mail: info-chr21@molgen.mpg.de * URL: http://chr21.fz-berlin.mpg.de/ AL163242: Submitted (10-Apr-2000).
FEATURES	Location/Qualifiers
SOURCE	1. 340000
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="21"
	<1. 75790
	/note="Accession No. AP001597"
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="21"
	/map="21q21.2"
SOURCE	

```
/clone="KB1951E10, 5' partial"
/clone_id="Keio BAC library"
737. .1092
/repeat_region
/note="HERV.L"
/rpt_family="LTR/Retroviral"
/rpt_type-DISPersed
complement(1763. .2230)
/note="MLT1D"
/rpt_family="LTR/MaLR"
/rpt_type-DISPersed
complement(3756. .4210)
/note="Charlie4a"
/rpt_family="DNA/MER1_type"
/rpt_type-DISPersed
4322. .4476
/note="L1P45"
/rpt_family="LINE/L1"
/rpt_type-DISPersed
5482. .3758
/note="AluDb"
/rpt_family="SINE/Alu"
/rpt_type-DISPersed
5759. .5866
/note="(GCAG)n"
/rpt_family="Simple_repeat"
/rpt_type-TANDEM
8386. .8557
/note="L2"
/rpt_family="LINE/L2"
/rpt_type-DISPersed
8822. .9116
/note="AluY"
/rpt_family="SINE/Alu"
/rpt_type-DISPersed
10343. .10460
/note="(TTCG)n"
/rpt_family="Simple_repeat"
/rpt_type-TANDEM
complement(10463. .10632)
/note="AluDb"
/rpt_family="SINE/Alu"
/rpt_type-DISPersed
complement(10635. .10669)
/note="Alu"
/rpt_family="SINE/Alu"
/rpt_type-DISPersed
complement(10852. .11010)
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type-DISPersed
11094. .11142
/note="(TG)n"
/rpt_family="Simple_repeat"
/rpt_type-TANDEM
complement(12123. .12810)
/note="LINE"
/rpt_family="LINE/L1"
/rpt_type-DISPersed
12952. .12985
/note="(TG)n"
/rpt_family="Simple_repeat"
/rpt_type-TANDEM
complement(13168. .13419)
/note="L1M4"
/rpt_family="LINE/L1"
/rpt_type-DISPersed
complement(13461. .13746)
/note="L1M4"
/rpt_family="LINE/L1"
/rpt_type-DISPersed
14398. .14604
/note="L1M2"
/rpt_family="LINE/L1"
```

```
/rpt_type-DISPersed
14604. .14753
/note="L1M4C"
/rpt_family="LINE/L1"
/rpt_type-DISPersed
14810. .15161
/note="L1"
/rpt_family="LINE/L1"
/rpt_type-DISPersed
complement(15167. .16789)
/note="MER52A"
/rpt_family="LTR/MER4_group"
/rpt_type-DISPersed
16799. .17007
/note="L1"
/rpt_family="LINE/L1"
/rpt_type-DISPersed
complement(17139. .17448)
/note="AluY"
/rpt_family="SINE/Alu"
/rpt_type-DISPersed
19049. .19312
/standard_name="D21S269"
/note="AFM263x15"
Accession No. Z17183"
19105. .19159
/note="(CA)n"
/rpt_family="Simple_repeat"
/rpt_type-TANDEM
complement(19358. .20123)
/note="L1M49"
/rpt_family="LINE/L1"
/rpt_type-DISPersed
complement(20142. .20413)
/note="L1"
/rpt_family="LINE/L1"
/rpt_type-DISPersed
complement(20415. .20792)
/note="MLT1A1"
/rpt_family="LTR/MaLR"
/rpt_type-DISPersed
20863. .20889
/note="(TC)n"
/rpt_family="Simple_repeat"
/rpt_type-TANDEM
complement(20970. .21075)
/note="L1"
/rpt_family="LINE/L1"
/rpt_type-DISPersed

repeat_region
complement(20970. .21075)
/note="L1"
/rpt_family="LINE/L1"
/rpt_type-DISPersed

Query Match 74.2% Score 985.6; DB 9; Length 340000;
Best local similarity 88.5%; Pred. No. 8.1e-289;
Matches 1120; Conservative 0; Mismatches 122; Indels 23; Gaps 4;

QY 1 tcaaatcgaagaagcttaagcttgcttaacccgcaaaagagggaagctgttatattt 60
|||||
Db 250910 TCACAAATGAGAGAGCTTTAGAAATGCTAATACATGAGAGAGGAGGAGAGCTTTATTTT 250969

QY 61 aggggaagaagctgctgtagtctgaatcaatccggaatcacttaactggaagttaaaga 120
|||||
Db 250970 AGGGGAAGAAGCTGCTTTATTTATGTTAATCAATTCGGAATCATCCAGAAAGTTAAAGA 251029

QY 121 aattgagatcgaatataatgtagagcagagagacctcaaaa-cactgcacctggggcc 179
|||||
Db 251030 AATTCAGATCGAATFACACGTAACAGAGAGAGCTTAAAAACAGTGCAGACCTGGGCCC 251089

QY 180 tcccaagcaatgagatgccctggagactccctctcctaagagccctgcagactataat 239
|||||
Db 251090 TCCACAGCAATGATGAGCCCGGATTCGCCCTCTTAAAGACCTGAGCAGCT--ATAT 251146

QY 240 tttaactcccttgagccctgtagtctcaactccctgtagttagcttgcctcccaaga 299
|||||
Db 251147 TTCTACTCCTCTTTGAGACCTGTATCTTTAACTCCGCTGTTAGTTGTCTCTTCAGAA 251206
```

OY	300	tttaagctgttaagctatacaaatagttcttcttaaatgtgaacccccaatgacgtacgtacata	359
Db	251207	TCGAAAGAGTAAACCTCAAAATCGTCTTTAAATATGGACCCCCAGATGACGATCATACATA	251266
OY	360	aactctaacgtgtagcccccctggaacgcgcctctgtagactatgctctgtatgtttaatgacatg	419
Db	251267	AGATCTACTGTAGAGACCCCTGGACACGCGCTCTAGGCCATGCTCCATATGTTAATGACATTGG	251326
OY	420	aagtcacccctcccgaggaaatctcaactgacacacccctactacacccccaattcaagtag	479
Db	251327	AAGGACCCCCCTCCAGAGAAATCTCAACATCTCACAAACCCCTACTACTCTCCAAATTCAGCAG	251386
OY	480	gaagcagcttagaagcagctgttctcaagccaaccccccacaagaactggtgttttccctgttag	539
Db	251387	GAACGACGTTACAGTGTGTCCTCTGGCCAACTCCCCAACAGCATTTGTAATTTCTCTGTGGG	251446
OY	540	agggtgtagctgagagaaagagactagctgtgaattctctagagctgactaagaatcccaagcc	599
Db	251447	AGGGGCGACTGAGAGACAGGACACTAGTGTGATTTCTTAGGCTGACTAGAAATCCTCTAAGCC	251506
OY	600	taactgtggaaggtgtaccgcatcattcttaacaatggtgctgtgcaactagctacaccc	659
Db	251507	TAGCTGGGAAGGTACCACTTCCACCTTTAAACAAGGGCTTTCACCTTAAGCTCACACCC	251566
OY	660	gaccacatc-----agagagctactaaatgtcttaactcagagcaaaaagagagtata	710
Db	251567	TACCAATTTGGATTAATAAGAGAGGCTCACTAAATGTCTAATTAAGGCATAAACAAGAGATA	251626
OY	711	agcaatagccaatcatctatctgctctgtagagacagcggaagagaaacaaagattgtgaata	770
Db	251627	AGAAATTAAGCCAAATCATCTATTCCTGTGAGACACAGCGGGAGGGAACAATGACAGATATA	251686
OY	771	aactcagggcatctaaagccagaaagcaaacaccccttgggttccctcccatgtatggga	830
Db	251687	AACCCAGGCATTTCCAGGCTCCCAAGGCAACGCCCCCTTTGGGTCCCTCTTTGTATAGGA	251746
OY	831	gctcctgtttcaactctatcttaactctataaatacattgcaatgtgcatctctgtgctgtg	890
Db	251747	GCTCTGT-----TTTCACGCTAATCAATCTTGCAACTGCTACTCTTCTGTGTCGTG	251796
OY	891	tttttatgtctcaagctgagctttgttctgcacatccacactgtctgtttgcacacgtata	950
Db	251797	TTTTTTTACGGCTCAAGGTGAGCTTTTCTCAACATCCACACACGCTGTTGGCCGCTTGG	251856
OY	951	cagacccgcgtctgacttcatccctcttggatccagagaaggtgtcgaactgtgtcctat	1010
Db	251857	CAGACCCGTGCTACTCTTCATCCCTCTCAAGTCTGGCAGGGTCTCCACTGTGTCTGTGAT	251916
OY	1011	ccagcgaagtiaaccattgycacactcccgatcagaagctaaaggctgycatgttttcctgcaag	1070
Db	251917	CCAGCAGAGGCACCATTTGCCACTCCGATCAGGCTTAAGGCTTGAAGCTTTCCTGACAA	251976
OY	1071	gctaagtgacctgggtttgttctctaataagaactgaacactgtgtcaactgtgttcattgtct	1130
Db	251977	GCTAAGTCCCTGGGTGCTGCTCTTAATCAAGGTGAACAATACTACTGCGGTTCATGTGTTCT	252036
OY	1131	cttccatgacccacagcgtcttcaatagagctataaactacacacgcgcatgagcccaagatcca	1190
Db	252037	CTTCCATGACCCATGGCTTCTTAATAGCTATTAACACTCACCCCATGCCCCAAGATTCCA	252096
OY	1191	ttctctgtgtatctgtgagcgaagaaccccaaggttcagagaangtgaagcttgccacatt	1250
Db	252097	TTCTCTTGAATTCGTGAGAGGCCAAGAACCCAGGTCAAGAAACAAGAGGCTGCCGCATCT	252156
OY	1251	ttggga 1255	
Db	252157	TGGAA 252161	

RESULT	13		
AP001600			
LOCUS	91097 bp	DNA	linear
AP001600			PRI 29-APR-2000

DEFINITION	Homo sapiens genomic DNA, chromosome 21, clone: KB45E1, APP-D21S292 region, complete sequence.
ACCESSION	AP001600
VERSION	AP001600.1
KEYWORDS	HTG.
SOURCE	Homo sapiens pre-pro-B cell cell_line: FLEB14-14 DNA, clone_11b: KeioBAC library clone: KB45E1.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (bases 1 to 91097)
TITLE	Shimizu, N., Kudoh, T. and Shibuya, K.
JOURNAL	Homo sapiens genomic DNA, chromosome 21, clone: KB45E1, APP-D21S292 region
REFERENCE	Published Only in Database (2000) In press
AUTHORS	2 (bases 1 to 91097)
TITLE	Shimizu, N., Kudoh, T. and Shibuya, K.
JOURNAL	Direct Submission
FEATURES	Submitted (04-APR-2000) to the DDBJ/EMBL/GenBank databases.
source	Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8587, Japan (E-mail: nshimizu@med.keio.ac.jp, Tel: 81-3-3351-2370, Fax: 81-3-3351-2370)
BASE COUNT	28309 a 16954 c 17419 g 28415 t
ORIGIN	Location/Qualifiers
	1..91097
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/cell_line="FLEB14-14"
	/cell_type="pre-pro-B cell"
	/chromosome="21"
	/clone="KB45E1"
	/clone_11b="Keio BAC library"
	Query Match 74.0%; Score 984; DB 9; Length 91097;
	Best Local Similarity 88.5%; Pred. No. 2.4e-288;
	Matches 1119; Conservative 0; Mismatches 123; Indels 23; Gaps 4;
QY	1 tcaaatccgaagagctttagacttcctaacccgcaaaaggaggagagctgtttat 60
Db	9856 TCAAAATCGAAGAGCTTTGAAATTCGTATATCTGAGAGAGGGGACCTTTTATTTT 9917
QY	61 aggggagaagaatgcgttagtagtcttaataatcctggaatccttactagagaag 120
Db	9918 AGGGAAGAGATCTGTATTATGTAAATCAATTGGAATCATCACCAAGAAAGTTAAAGA 9977
QY	121 aattgagatcgaaatataatgtagagcagagagacttcaaa-cactgacccctgggccc 179
Db	9978 AATTCAAGTTCAAATPACAACGTAGAACAGAGAGCTTAAAAAACACTGGACCTGGGGCC 10037
QY	180 tccctagccaatgagatgcctctgagactctccctctttagagactctagagctataat 239
Db	10038 TCCTCAGCCAAATGAGATGCCCTGATTTCTCCCTTTAGAGACTCTAGACGCT---ATTAT 10094
QY	240 tttaactccctcttggagccctgtagtcttcctaactctccctgtgtaagttgtctctccaga 299
Db	10095 TTCTACTCTCTCTTTGGACCCCTTATCTTAACTCCGCTTAAGTTTGTCTTCCAGAA 10154
QY	300 ttgaagctgttaagctacaatagttcttcaatiggaaccgccagatgscagttcatgacta 359
Db	10155 TCGAAGATGTAAATCTACAAATGTTCTTCAATATGGACCCCGAGATGCGATCCATGACTA 10214
QY	360 aaatctaccgttggagccctcgagccggcgcgtctagagactagctgtgtaagttacatg 419
Db	10215 ACATCTACTAGGAGACCCCTGGAGCAGCGGGCTAGGCCATGCTCCAAATGTTAATGACATGG 10274
QY	420 aagtaacccctccgagagaatctcaactgtgcaaacccctactacaactccaatcaagtag 479
Db	10275 AAGGCAACCCCTCCCAAGGAATCTCAACATGCAACACCCCTACTATGCTCCAAATTCAGCAG 10334
QY	480 gaagcaagtttagagcaatttcaagccaactctcccaacatacttggqtttctctgttag 539

```

Db 10335 GAACGAGTTACAGTGTCTCTGGCCAACTCCCAACGACATTTGTATTTCTCTTGGG 10394
Qy 540 agggctgagctgagagacagactagctgcttctcttaagctgactaaagcccccnaagcc 599
Db 10395 AGGGGCGACTAGAGACAGGACTAGCTGATTTCTTACGGCTGACTGACATTCCTTAAGCC 10454
Qy 600 tancctggaaggtgacgcgacatcctttaacatgagggcttggaacttagctacaccc 659
Db 10455 TAGCTGGGAAGGTGACCACTTCCACCTTTAAACAGAGGCTTGCAACTTACCTCACACCC 10514
Qy 660 gaccacatc-----agagaagctcaactaaatgcttaacgaagcaaaagagagtaaa 710
Db 10515 TACCAATTGGATAGTAAAGAGAGTCACTAAATGCTATTAGGCAAAAACAGAGAGTAA 10574
Qy 711 agacatagcaccatcatctatctgctgagagacagcgaggaagcaagagtgatata 770
Db 10575 AGAATATACCAATCTTCATCTTGCCTGAGACGACAGCGGAGGACATGACAGGATATA 10634
Qy 771 aactcagggcatcacaagcagaacagacaccccttggctccctcccatgtatgagga 830
Db 10635 AACCCAGGCAATTCAGACCTGCGACGAGCAACCCCTTTGGTCCCTCTCTTGTATGGGA 10694
Qy 831 gctctgtttcaccctctatcttaactctataaatacgaactgcaactcttctgttcgtg 890
Db 10695 GCTCTGT-----TTTCACTCTATTCAATCTTGCACCTGCACTCTTGTGTCGCTG 10744
Qy 891 tttttatgctcacaagctgagcttctgtgcacatccacactgctgttctgacacgtca 950
Db 10745 TTTGTATAGGCTCAAGCTGAGCTTTTGTGCTACATCCACGACACTGCTGTTCGCCCGTGG 10804
Qy 951 cagacccgctgctgactcactcacccttgatcagacagagatgctccactgctgctcgat 1010
Db 10805 CAGACCCCTCTCTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10864
Qy 1011 ccagcagagtgacccatgacccatgacccatgacccatgacccatgacccatgacccatg 1070
Db 10865 CCAGCAGAGGCAACCATTCCTCCCTCCGATCCAGCTTAAGGCTTGTGCTATGCTGCACA 10924
Qy 1071 gctaaagctgctgaggttctgcttaataagactgaacactgctgaggttctgctgcttct 1130
Db 10925 GCTAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10984
Qy 1131 ctccaagacacacagcgtctcctaataagactgaacactgacccgcatggtcccaagatcca 1130
Db 10985 CTTTCATGACACCATGCTCTTAATAGAGCTTAACACCTCACCGCATGCGCCCAATTCACA 11044
Qy 1191 ttcctgtgatactgtgagagcagaagacccagatcagaagaangtgaagcctgacacatc 1250
Db 11045 TTTCCTGGATTCGCTGAGGCGCAAGACCCCGATCAGAGACAGAGGCTGCCGCATCT 11104
Qy 1251 tggga 1255
Db 11105 TGGAA 11109

```

```

RESULT 14
AC008121 172721 bp DNA linear HTG 16-OCT-2001
LOCUS Homo sapiens chromosome 12 clone RP11-407N8, WORKING DRAFT
DEFINITION AC008121.25 GI:14190616
ACCESSION AC008121.25
VERSION AC008121.25
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 172721)
AUTHORS Muzny,D.M., Adams,C., Adio-Ogunola,B., All-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,

```

TITLE JOURNAL AUTHORS REFERENCE JOURNAL

COMMENT

```

Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gilly, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlik, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,
Hollins, B., Homel, F., Howard, S., Huber, J., Hulik, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W.,
Louisege, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapa, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M.,
Mel, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabhat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S.,
Oguni, M., Okunola, G., Oragunye, N., Oriedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojudo, K., Rolfe, M.,
Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N.,
Stinson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Swalek, A., Taber, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Vinson, R.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Vellard, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G., and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 172721)
Worley, K.C.
Direct Submission
Submitted (24-JUL-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 23, 2001 this sequence version replaced gi:13940575.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMR
Center clone name: RP11-407N8
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye 41% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 152097 bases at least Q40
Consensus quality: 166730 bases at least Q30
Consensus quality: 173153 bases at least Q20
Estimated insert size: 173988; sum-of-ctrls estimation
Quality coverage: 0x in Q20 bases; agrose-fp estimation
Quality coverage: 3x in Q20 bases; sum-of-ctrls estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

```


RESULT 15
AC016173/c
LOCUS
DEFINITION Homo sapiens chromosome 2 clone RP11-440C21 map 2. WORKING DRAFT
AC016173
AC016173
AC016173.2 GI:7657740
VERSION
KEYWORDS
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
2 (bases 1 to 145785)
Birten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,D., Barna,N., Beckert,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeRubeis,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Goto,Y.,
Gallagan,J., Gardyna,S., Grant,G., Hagos,B., Heathford,A., Horton,L.,
Howard,J.C., Johnson,R., Jones,C., Kann,L., Karalas,A., Klein,J.,
Lemock,J., Liu,C., Locke,K., Macdonald,P., Marquis,N.,
McManis,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Teefaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyma,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 27, 2000 this sequence version replaced gi:6466621.
All repeats were identified using RepeatMasker:
Smt, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3813
Center clone name: 440_C-21
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 136152 bases at least Q40
Consensus quality: 140073 bases at least Q30
Consensus quality: 141708 bases at least Q20
Insert size: 18700; agarose-fp
Insert size: 144185; sum-of-ctnigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-ctnigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 ctnigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the ctnigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1331: ctnig of 1331 bp in length
* 1332 1431: gap of 100 bp
* 1432 2628: ctnig of 1197 bp in length
* 2629 2728: gap of 100 bp
* 2729 3745: ctnig of 1017 bp in length
* 3746 3845: gap of 100 bp

FEATURES
source
* 3846 6327: ctnig of 2482 bp in length
* 6328 6427: gap of 100 bp
* 6428 7950: ctnig of 1523 bp in length
* 7951 8050: gap of 100 bp
* 8051 10578: ctnig of 2528 bp in length
* 10579 10678: gap of 100 bp
* 10679 15308: ctnig of 4630 bp in length
* 15309 15408: gap of 100 bp
* 15409 20910: ctnig of 5502 bp in length
* 20911 21010: gap of 100 bp
* 21011 25032: ctnig of 4022 bp in length
* 25033 25132: gap of 100 bp
* 25133 29129: ctnig of 3997 bp in length
* 29130 29229: gap of 100 bp
* 29230 36550: ctnig of 7321 bp in length
* 36551 36650: gap of 100 bp
* 36651 48963: ctnig of 12313 bp in length
* 48964 49063: gap of 100 bp
* 49064 62673: ctnig of 13610 bp in length
* 62674 62773: gap of 100 bp
* 62774 75359: ctnig of 12586 bp in length
* 75360 75459: gap of 100 bp
* 75460 90704: ctnig of 15245 bp in length
* 90705 90804: gap of 100 bp
* 90805 109887: ctnig of 19083 bp in length
* 109888 109987: gap of 100 bp
* 109988 145785: ctnig of 35798 bp in length.
Location/Qualifiers
1. 145785
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-440C21"
/clone.lib="RPC1-11 Human Male BAC"
1. 1331
/note="assembly_fragment"
1432. 2628
/note="assembly_fragment"
2729. 3745
/note="assembly_fragment"
3846. 6327
/note="assembly_fragment"
clone_end:17
vector_side:right"
6428. 7950
/note="assembly_fragment"
8051. 10578
/note="assembly_fragment"
10679. 15308
/note="assembly_fragment"
15409. 20910
/note="assembly_fragment"
21011. 25032
/note="assembly_fragment"
25133. 29129
/note="assembly_fragment"
29230. 36550
/note="assembly_fragment"
36651. 48963
/note="assembly_fragment"
49064. 62673
/note="assembly_fragment"
62774. 75359
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
75460. 90704
/note="assembly_fragment"
90805. 109887
/note="assembly_fragment"
109988. 145785
/note="assembly_fragment"

BASE COUNT 4438 a 27590 c 27266 g 44991 t 1600 others
 ORIGIN

Query Match 73.7%; Score 979.2; DB 2; Length 145785;
 Best Local Similarity 88.2%; Pred. No. 7.2e-287;
 Matches 1116; Conservative 0; Mismatches 126; Indels 23; Gaps 4;

QY 1 tcaaatcgaagagcttgaacttgcctaacccgcaaaagggggggaacccgtttattttt 60
 DB 7898 TCAAAATCGAAGACCTTTAAATTTGCTAAATCACTGAGAGAGGGGAACTTTTATTTT 7839
 QY 61 agggagaagatgctgttagatgttaatcaatcctggaatcaatcactlgaagaagttaaaga 120
 DB 7838 AGGGGAAGATGCTGTATATGTATATCAATTGGAATCATCAACCAAGAAAGTTAAGA 7779
 QY 121 aatttgaatcgaataataatgttagagcagagacccctcaaaa-caatgcaccttgggccc 179
 DB 7778 AATTCAAGATCGAATAACAGTAGAACGAGGAGGCTTAAAAAACCTGAGACCTGGGGCC 7719
 QY 180 tccctcaagcaatgagatgacctgacctccctccttagagaccttagcagcgtataat 239
 DB 7718 TCCTCAAGCAATGAGATGCTGATTTCCCTTTCTTGAAGCTCTGACAGCT--ATAT 7662
 QY 240 tttaactcctcttggacctgtatcttcaactccttctgttaagtctgtctctccagaa 299
 DB 7661 TTCTACTCCCTTTGGACCTGTATCTTTAACCTCCGCTGTAAAGTTGTCTCTCCAGAA 7602
 QY 300 ttgaagctgtaagctacaatagttctcaaatgaaacccagatgacgtacatgaacta 359
 DB 7601 TCGAAGATGTAAACTCAAAATGCTCTCAAAATGAGACCCCAAGATGACATGACTA 7542
 QY 360 aaatcaccgtgagacccgtgacgtctgaactatgctcgtatgtaataatgactg 419
 DB 7541 AGATCTACTAGAGACCCCTGAGACCGCTGACCCATGCTCAATGTATATGACATG 7482
 QY 420 aagtcacccctcccgaggaatctcaactgacacacccctactacactccaatcagtag 479
 DB 7481 AAGGCACCCCTCCCAAGGAATCTCAACTGACACACCCCTACTATGCTCCAAATTCAGCAG 7422
 QY 480 gaagcagttgaagcagttgtcgaagcaaacctcccaacagtaacttgggttctcgttgag 539
 DB 7421 GAAGCAGTTGACAGTGTCTCGGCAACCTCCCAACAGCATTTGTATTTTCCGTGGG 7362
 QY 540 aggtgagcagagagacagactgagtttccctagagctgaactgaactgaactcacaacc 599
 DB 7361 AGGGGGCACTGAGAGACAGACTGAGATTTCTTAGCTGACTGAGAAATCCCTAAAGCC 7302
 QY 600 tancctggaaagtgaccgcacalccttaaacatggggttgcacttgaacttgaactcacc 659
 DB 7301 TAGCTGGGAAGGTGACCACTTCCACTTTAAACAGAGGGCTTGCACCTTAGCTCACACC 7242
 QY 660 gacccaatc-----agagaactcaataatgtcctaatacagcaaaaacaggaagtaa 710
 DB 7241 TACCAATTGATAGTAAGAGAGAGCTCAATAATGCTAATTAAGCAAAACAGAGAGTAA 7182
 QY 711 agcaatagcaatcatatattgcttgaagcagcagcggaaggaaggaatgtggatata 770
 DB 7181 AGAAATAGCCATATCATTTGCTGAGAGCACAGCGGAGGACAAATGACCAAGATATA 7122
 QY 771 aactcaggaatcgaagcagcaagaacaccccttgggtccctcccatgtatgga 830
 DB 7121 AACCAAGGCAATTCAGGCTGAAAGGCAACCCCTTTGGGTCCCTCTTTGTATGGGA 7062
 QY 831 gctctgtttcaactctatctcaacttaataatgaactgaactccttctgtccgtg 890
 DB 7061 GCTCTGT-----TTTCACTCTATTCATCTTGCACACTGCACTTCTGTGTCGTG 7012
 QY 891 ttttttgaagcagcgttcttctgcaatcgaacacactgtcttgcacacgtca 950
 DB 7011 TTTGTATAGGCTCAAGCTGAGCTTTTGTGCTCACCACTCACTGCTGTGTCGCGCGTGG 6952
 QY 951 cagaccgctgtgacttccacttggatccagcagagtgctcactgtcctcgtat 1010

DB 6951 CAGACCCGCTGCTGACTTCCATCCCTCCAGATGCGCAGGGGTGNCACATGTGCTGAT 6892
 QY 1011 ccagcagagtaaccatctgcacactcccgatacagcgttaagcctgtgcactgtctcgtatg 1070
 DB 6891 CCAGCGAGGACCCCATTTGCGCTCCCAATCGGGCTTAAAGCCTTGCATTTGTTCTGCACA 6832
 QY 1071 gctaagtcgctgtgtgttctcctaataagactgaacactgttgaactgtgttcacatgtctc 1130
 DB 6831 GCTAAGTCCCTGGGTTGCTCTTAATCAAGCTGAACACTAGTCACTGGGTTCCATGTTCT 6772
 QY 1131 ctccatgacccacgcttcttaataagactataaacactacccgaatgcccgaatcca 1190
 DB 6771 CTTCATGACCCCATGCTTCTTAATAGAGCTATTAACACTCAACCGCATGGCCCAAGATTCCA 6712
 QY 1191 ttcctgtgactgtgtgagcgaagaacccagatcgaagaangtgaagctgtccacatc 1250
 DB 6711 TTCTTGGAATCCGTGAGGCCAAGAACCCAGTCAAGAAACAGAGGCTGTGCTCATCT 6652
 QY 1251 tgaga 1255
 DB 6651 TGGAA 6647

Search completed: June 20, 2002, 10:15:40
 Job time: 12665 sec

THIS PAGE BLANK (USPTO)